Cooking fuel choice and child mortality in India*

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

This version: April 9, 2024

Abstract

How serious is indoor air pollution (IAP) as a threat to infants and children? This paper estimates the impact of cooking fuel choice – a predominant source of IAP – on under-five mortality in India, where reliance on biomass fuels such as firewood, animal dung, and agricultural waste is pervasive. Leveraging forest cover and agricultural land ownership for identification and nationally representative data, we find that solid fuel use for cooking significantly increases the child mortality rate - mainly driven by neonatal mortality in the first 28 days after birth. The mortality effect is higher for girls than boys and is magnified in relatively small households where there is limited scope for the division of labor between childcare and cooking responsibilities. Among polluting fuels, we find that biomass fuels drive the impact of polluting fuel use on child mortality.

JEL classification: I12, J13, O15, Q52
Keywords: Energy access, Indoor air pollution, Infant health, Heterogeneous impact

---

*We thank Ali Abbas, Chris Barrett, Dusan Drabik, Andrew Foster, Maulik Jagnani, Gary Lin, Anh Nguyen, Hemant Pullabhotla, Avralt-Od Purevjav, participants at the 2019 Northeastern Universities Development Economics Consortium (NEUDC) Conference at Northwestern University; 2021 International Economic Association (IEA) virtual World Congress, Indonesia; 2023 Conference on Economic Theory and Policy, Presidency University, Kolkata, India; 2024 Annual Economics Conference, Ahmedabad University, India; Wageningen University, NL; two anonymous referees and the Associate Editor of this Journal for their comments and suggestions. An earlier version of this paper was circulated as "Cooking Fuel Choice, Indoor Air Quality and Child Mortality in India". The usual disclaimer applies.

†Basu: Charles H. Dyson School of Applied Economics and Management, Cornell University, 441 Warren Hall, Ithaca, NY 14853, United States & Institute for the Study of Labor (IZA), Bonn, Germany (email: arnab.basu@cornell.edu); Byambasuren: Charles H. Dyson School of Applied Economics and Management, Cornell University, 438 Warren Hall, Ithaca, NY 14853, United States (email: tb497@cornell.edu); Chau: Charles H. Dyson School of Applied Economics and Management, Cornell University, 201A Warren Hall, Ithaca, NY 14853, United States & Institute for the Study of Labor (IZA), Bonn, Germany (email: hyc3@cornell.edu); Khanna: Department of Economics and Environmental Studies Program, Binghamton University, Library Tower 1004, Binghamton, NY 13902-6000, United States (email: nkhanna@binghamton.edu).

1 Introduction

Indoor air pollution (IAP) is the leading environmental factor behind mortality in India, accounting for about 40% of the 1.2 million deaths in 2017 (Global Burden of Disease 2017). Globally, it leads to around 3.8 million premature deaths per year - a number that far exceeds mortality attributed to malaria and tuberculosis combined (WHO, 2021). This pollution is caused by using biomass fuels (firewood, animal dung, and crop waste), coal, and kerosene for cooking and heating. Despite several initiatives to encourage the adoption and use of cleaner fuels and improved cookstoves in various countries, 36% of the global population continues to use polluting cooking fuels (Stoner et al., 2021).

This paper explores an important but under-researched dimension of the cost of polluting cooking fuel use in developing countries: the mortality of young children. In many developing countries, women disproportionately and simultaneously shoulder the burden of cooking and childcare. In India, the setting of our study, 56% of children under age 5 always stay with their mothers, including during cooking (Rehfuess et al., 2011; Edwards and Langpap, 2012). Young children are thus particularly vulnerable to the adverse health risks associated with indoor air pollution (IAP) in households that use polluting cooking fuels.

While the link between improved cookstove adoption and improvements in health outcomes has been explored in a few studies (Diaz et al., 2007; Smith-Sivertsen et al., 2009; Hanna et al., 2016; Barron and Torero, 2017), to the best of our knowledge, causal estimates of the direct cost of using biomass fuels for cooking on child mortality are non-existent. To this end, we leverage a large-scale household survey in India – the Demographic and Health Survey (DHS) – which contains individual- and household-level health and demographic information as well as details on the fuels used for cooking between 1992 and 2016. Our data offers several unique features that ensure the external validity of our findings and generate new insights on the impact of biomass fuel use on infant mortality. First, the DHS contains data on 10 different types of fuels used by Indian households such as kerosene, LPG, and biomass. Thus, compared to studies that utilize a restrictive sample to analyze the effectiveness of cookstove adoption or the switch from kerosene to LPG as cooking fuel choices, our study exploits a wider range of cooking fuels. Second, the DHS reports a complete record of births and deaths, allowing us to extensively analyze the environmental cost of infant mortality across a wide range of age groups. Specifically, this enables us to examine heterogeniety in the IAP associated mortality risk by sorting children into five distinct age groups: neonatal (0-28 days), post-neonatal (28 days-1 year), infant (0-1 year), child (1-5 years), and under-five (0-5 years). Third, the DHS collects data from both urban and rural areas, and this unique feature allows us to provide broadly representative empirical estimates of the causal relationship between cooking fuel choice and the mortality risks via exposure to IAP faced by children.

Establishing a causal link between IAP and under-five mortality, however, is challenging since the choice of cooking fuel may well be endogenous to the incidence of mortality within a household due to omitted variables and simultaneity biases.[1] In terms of simultaneity bias, households can be caught in a self-perpetuating cycle wherein they are only able to afford cheaper and more polluting cooking fuel options, which adversely affects household health and earnings, perpetuating the use of polluting fuels (Hanna and Oliva, 2015; Graff Zivin and Neidell, 2012, 2018; Chang et al., 2016, 2019). In terms of omitted variables bias, unobserved characteristics might confound with cooking fuel choice. To deal with these identification threats, we leverage two instrumental variables for the fuel choice of a household – forest cover and agricultural land ownership.

The causal relationship between cooking fuel choice and child mortality relies on the choice and validity of the instruments. Thus, at the onset, it is important to emphasize that most of the children in the DHS surveys are from rural areas of India, and a large number of households are small agricultural landholders. Given that household use of biomass fuels – especially crop residues, animal waste, and firewood – is highly correlated with farming, livestock ownership, and easier access and/or a low opportunity cost of firewood use, we choose agricultural landownership and density of forest cover within a district as our instruments. We subsequently undertake a set of rigorous tests to check for the validity of these two instruments. In addition to the tests of relevance (instruments are correlated with the endogenous regressor) and independence (instruments are uncorrelated with any confounders of exposure-outcome relationship), we test the exclusion restriction (instruments affect the outcome only through the endogenous regressors) via a zero-first-stage test. Under this test, the reduced-form effect of the instrument on the outcome is plausibly zero in a subsample for which the effect of the instrument on the treatment variable is zero (Van Kippersluis and Rietveld, 2018). As we discuss in detail later, we use a subsample of small households with three or fewer members from urban areas as the zero-first-stage sample since these households have limited access to forests as a source of firewood and limited capability of growing crops and using agricultural crop residues for cooking. We also test the exclusion restriction by analyzing whether our instruments are associated with other child characteristics and maternal health-seeking activities and find that the instruments are not relevant to these outcomes. Finally, we examine whether under-five child mortality can be explained by either household size or the mother's educational attainment as an additional test of the exclusion restriction. We find that neither of these two variables instrumented by agricultural land ownership status and forest cover affects under-five mortality, unlike cooking fuel choice.

We thus estimate the causal effect of IAP, measured by primary cooking fuel use, on under-five and infant mortality in a large-scale setting, relying on plausibly exogenous variations

[1]There is a robust parallel literature on household 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). These studies provide evidence that households react to changes in health outcomes due to outdoor air pollution by adjusting their behavior, adopting new technologies, and investing in protective goods in response to health shocks.

in IAP introduced by the forest cover and agricultural land ownership status. The results suggest highly heterogeneous local mortality effects by children's age and gender and household characteristics. Specifically, we find that the effect of polluting cooking fuel use on under-five and infant mortality rates is 0.040 (standard error = 0.020) and 0.041 (standard error = 0.019). This translates into 27 under-five children and infants per 1,000 live births that would have been saved economy-wide if all households used clean fuels for cooking.[2] Moreover, heterogeneous treatment effects by the child's age and household size show that the risk of child mortality due to cooking fuel choice is driven by mortality within the first 28 days of birth, and it is the highest in households with five or six members. Specifically, the estimated effect on neonatal mortality rate is 0.030 (standard error = 0.015), implying the loss of 20 neonates for every 1,000 live births that would have been saved if all households used clean fuels for cooking. We also document that the adverse effects of polluting fuel use on infant and under-five mortality are concentrated on girls, with no impact on boys, which is consistent with a related literature that suggests a bias in the mistreatment of young girls (Anderson and Ray, 2010).[3] Our estimates are also robust to several alternative specifications, such as a more granular definition of polluting cooking fuels (the effect of fuel dirtiness on under-five and infant mortality rate is 0.007, and the impact on neonatal mortality rate is 0.005), and use of satellite versus Census-based forest cover data from a different period than the one used for the baseline analysis.

Our paper is related to an emergent literature that explores the link between IAP and health outcomes in developing countries.[4,5] This literature can be divided into two groups along methodological lines: one that uses experimental approaches like randomized control trials (RCTs) to evaluate the health benefits of choosing cleaner fuels for cooking and the other that uses quasi-experimental methods to causally estimate the fuel choice and health nexus, e.g., via difference-in-difference (DID) and instrumental variable (IV) designs. Table 1 summarizes the context, identification strategies, and findings from the literature.

RCTs analyzing the health impacts of improved cooking stoves arrive at orthogonal conclusions. Diaz et al. (2007) and Smith-Sivertsen et al. (2009) find that improved cooking stoves reduce headaches, sore eyes, and chronic respiratory symptoms like wheezing for women during 18 months of their adoption in San Marcos district, Guatemala. However, Hanna et al.

[2]More precisely, our estimates imply that 26.90 under-five children and 27.46 infants would have been saved per 1,000 live births if all households used clean fuels for cooking.

[3]This finding is also consistent with the fact that infant mortality is higher for girls than boys in the early ages (`https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health`).

[4]There is also rich literature on the relationship between cooking fuel choice and non-health outcomes, such as women empowerment and time use, in developing countries. For example, Lee et al. (2023) show that shifting from biomass to clean fuels is associated with women empowerment, which also endogenously determines the cooking fuel choice in India (Ghosh et al., 2024).

[5]A complementary literature focuses on outdoor air quality, exploiting 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).

(2016) find that improved cookstoves did not affect various health outcomes, including sore eyes, headache, wheezing, and other respiratory symptoms among primary cooks (women) and children in the household in rural Orissa, India. Smoke exposure fell in the early years of installation, but the smoke-reducing impact of improved stoves disappeared after two years due to improper maintenance. Other studies that focus on the effectiveness of specific policies and programs (e.g., improved cooking stoves, house construction, and voucher allocation for electrification) conclude that switching to cleaner cooking fuels is associated with better health outcomes. For example, Barron and Torero (2017) suggest that electrification program in northern El Salvador reduces the acute respiratory infections among children under six.

Studies using a quasi-experimental DID strategy include Imelda (2018, 2020) who estimates a significant reduction in the infant mortality rate resulting from an Indonesian government program of subsidizing households to switch from kerosene to LPG for cooking. Silwal and McKay (2015) show that firewood use for cooking harms an individual's lung capacity by 9.4 percent in Indonesia using proximity to the nearest market as an instrument. Edwards and Langpap (2012) find that firewood use in Guatemala, instrumented via household's ownership of gas stoves, increases the probability of under-five children attached to mothers who cook experiencing a respiratory infection. Pitt et al. (2006) find that four hours spent cooking per day increases the likelihood of adults' respiratory symptoms by 10.8 percentage points and proximity to stoves adversely affects the respiratory health of women and young children in both rural Bangladesh and rural India using gender-specific hierarchies as an instrument. Finally, Liu et al. (2020) suggest that non-solid fuel use benefits an elder's ability to cope with daily activities (eating, bathing, housekeeping, financial and medical management) in rural China using the share of clean fuel users in the village as an instrument.

This paper advances prior work in two ways. We provide population-based evidence on the effect of cooking fuel choice on infant mortality, a topic that has not been addressed by prior population-based studies that use the IV approach. Furthermore, we leverage a new identification strategy based on variation in household cooking fuel choice in response to environmental and household factors, including forest cover and agricultural land ownership. We thus depart from prior works that have focused on the effects of specific programs or experiments and broaden the outcomes covered beyond respiratory diseases.

The rest of the paper is structured as follows. Section 2 describes the data and empirical approach. Section 3 details our identification strategy, and Section 4 presents estimation results. Section 5 checks the robustness of our results, while Section 6 concludes.

2 Data and empirical strategy

Our empirical analysis utilizes three datasets. The first, with nearly 0.4 million observations, is the nationally representative Demographic and Health Surveys (DHS) in India. The DHS collects individual-level data on mortality incidence and other socio-economic characteristics for every member residing within the sample household. It also contains household-level information on wealth, type of housing, location of the residence, and agricultural land ownership status. Importantly, for our analysis, DHS includes information on the specific type of cooking fuel that a household uses, allowing us to approximate variations in indoor air quality at the household level. To date, four rounds of the survey have been conducted since 1992–93.[6] We rely on three rounds of this survey: DHS-1 (1992–93), DHS-2 (1998–99), and DHS-4 (2015–16)[7] which includes a total of 268,819 ever-married women of reproductive ages between 15–49 years (69,107 from urban and 199,712 from rural areas)[8] and a pooled dataset of 369,416 singleton live-born children, of whom 19,474 died in the 5-years preceding the respective survey year.

To address the endogeneity between child mortality and a household's choice of cooking fuel we use two instrumental variables that proxy the latter: land ownership and forest cover at the district level, both of which eases a household's access to crop waste and wood for cooking fuel purposes. While the data on land ownership is available from the DHS surveys, we use satellite data on forests from the Planning Commission of India which serves as the second data source in our analysis and the primary database on land use in the country. Lastly, and as a robustness check, we also obtain surface area of land covered by forests at the village and city block level from the 2011 Census of India.

To investigate the causal effect of cooking fuel choice on the mortality rates of children, we specify the following reduced form relationship:

$$\begin{aligned} \textit{Child Mortality}_{ihdst} = \alpha + \beta \textit{Polluting Fuel}_{hdst} + \textit{Household}_{hdst}\boldsymbol{\gamma} + \textit{Mother}_{ihdst}\boldsymbol{\lambda} + \\ + \textit{Child}_{ihdst}\boldsymbol{\delta} + \textit{District}_{dst}\boldsymbol{\pi} + \eta_{st} + \varepsilon_{ihdst}, \end{aligned} \tag{1}$$

where $\textit{Child Mortality}_{ihdst}$ is one of the five age-specific binary variables measuring mortality: (i) under-five (younger than five years old), (ii) child (toddlers and children between the ages of one and five), (iii) infant (younger than one-year-old), (iv) post-neonatal (infants between 28 days and one year old), and (v) neonatal (first 28 days of life). $\textit{Child Mortality}_{ihdst}$ is assigned a value of 1

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

[7]We exclude DHS-3 (2005–06) in our empirical analysis due to the absence of district identifiers in the questionnaire in this round for HIV testing confidentiality.

[8]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.

if the mortality happened over the given age periods, and 0 if the child survived during the age-period for child i, in household h, in district d of state s, in survey year t. While our primary outcome variable is under-five mortality, it is important to acknowledge that the detrimental effects of pollution can be age-specific with exposure to IAP possibly more detrimental for younger children.

The key regressor is a binary variable ($\textit{Polluting Fuel}_{hdst}$) which indicates the primary cooking fuel type and is hence a proxy for indoor air pollution within a household. Ten types of cooking fuel are reported in the DHS datasets, and we classify these fuels into two groups – clean and polluting – based on the level of smoke produced during cooking. 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.[9,10] The vectors $\textit{Household}_{hdst}$, $\textit{Mother}_{ihdst}$, $\textit{Child}_{ihdst}$, and $\textit{District}_{dst}$ are respectively household-specific (h) characteristics, mother-specific characteristics for child i, child-specific (i) characteristics, and district-specific (d) characteristics that are potential determinants of under-five mortality. We discuss each of these in detail below.

Household characteristics. We include the urban or rural classification for place of residence, household wealth index (high wealth, middle wealth, or low wealth)[11], type of house: either pucca (made with durable materials) or semi-pucca / kachha (constructed with non-durable materials like mud, cloth or unprocessed wood), and number of household members[12] as potential socio-economic factors (Bassani et al., 2010; Ezeh et al., 2014; Naz et al., 2016).

A significant factor determining the link between cooking fuel choice and the intensity of IAP is the location within the household where food is cooked. We exploit detailed observations on

[9]Empirical evidence suggests that households often use multiple fuels simultaneously, as in the energy ladder model (Heltberg, 2005) or the alternative, fuel stacking (Heltberg et al., 2000). However, since the DHS survey records only the primary fuel used for cooking, we are not able to address the issue of multiple fuel use, and focus instead on assessing the implications of primary fuel use.

[10]As an alternative to our baseline dummy variable of polluting fuel use, we also considered access to electricity as another measure of cooking energy source to capture fuel stacking behavior. However, as we show in Table B.1, access to electricity is not correlated with the use of electricity as a primary cooking fuel source – the relationship is almost negligible when no controls are included (Column (1) of Panel A), and zero once we control for household characteristics in our preferred specification (Column (4) of Panel A). Electricity access is positively and statistically correlated with the use of electric fans and television at the 1% level, even after controlling for fixed effects and household characteristics (Panels B and C). We conclude that access to electricity is not a sufficient condition for the use of electricity as a primary cooking energy source.

[11]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", "middle wealth", and "high wealth" are defined as the lowest 40% of households, the middle 40% of households, and the top 20% of households, respectively (Filmer and Pritchett, 2001).

[12]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. The average household in the survey has seven members. However, as many as 46 people can be present in one household (the maximum is 41 in the DHS-4 data), and in general, households with greater than 25 members could be considered as outliers. A noteworthy correlation between household size and cooking fuel choice stems from the fact that manufactured gas stoves limit the volume of food that can be cooked in a single sitting relative to wood-burning furnaces that can be tailored to any size.

cooking location — in the living room, in the kitchen separate from the living room, in a separate building, or outdoors — embedded within the DHS-4 survey to proxy for the extent of exposure to polluting fuels.[13]

Mother characteristics. The key characteristics of the mother that correlate with infant health and mortality are educational attainment (secondary/higher, primary, or no education) which can signal knowledge about child care and about available health interventions (vaccines, supplemental nutrition, etc.), and mother's age (<20, 20-29, 30-39, and 40-49 years) that proxies child-rearing experience.[14]

Child characteristics. We consider the gender of the child and breastfeeding status (whether the child was ever breastfed): breastfeeding may protect children from under-five mortality, particularly in the neonatal and post-natal periods (Cushing et al., 1998; Arifeen et al., 2001; Black et al., 2003; Ezeh et al., 2014).

District characteristics. Biomass fuel use contributes to both indoor and outdoor air pollution and infant mortality is affected by both. Potentially, one can either use district-wise concentrations of fine particulate matter ($PM_{2.5}$—particulate matter or inhalable particles with diameters of 2.5 micrometers and smaller) or adult mortality rate to approximate outdoor air pollution since adults tend to be more exposed to ambient air pollution which is likely to contribute to reduced life expectancy via stroke, heart disease, lung cancer and respiratory diseases (Chen et al., 2013; Ebenstein et al., 2017). A comparison of India's 2014 district-wise ambient air pollution ($PM_{2.5}$) with the 2015-2016 district-level adult mortality rate in Figure B.1 shows that these two indicators are positively associated. Furthermore, the adult mortality measures in our sample are positively correlated to $PM_{2.5}$ level at the 1 percent level. In particular, the mortality rate for 15-49 years old adults is weakly but positively correlated to ambient air pollution, while the correlation is much stronger for mortality of adults older than 50 years, which is consistent with estimates from Global Burden of Disease (2017).[15] We use the adult mortality

[13]The DHS-1 and DHS-2 questionnaires ask respondents whether the household has separate room as the kitchen. This does not allow us to distinguish between cooking in different indoor locations and outdoor cooking. The DHS-4 (2015-16) includes a variable indicating whether a household cooks inside the house, in a separate building, or outdoors in addition to a question of the separate kitchen. We combine these two variables to construct the cooking location measure used in our regressions.

[14]The care burden that mothers carry can also increase with employment. While the DHS data reports information on the mother's current working status, employment in the last 12 months, and other related employment-related variables such as type of earnings from work, nearly 85% of the observations of these variables are missing in the DHS-4 data. Since this is the primary dataset in our IV estimation, we do not include the mother's employment status or other related variables in our analysis to avoid potential bias from sample selection. However, we indirectly account for the mother's employment status since we control for the mother's characteristics correlated with her employment, such as the mother's educational level and household income, as shown in Afridi et al. (2018) and Sarkar et al. (2019). Previous evidence from India suggests that women with higher education spend more time on domestic work and childcare and less on market work (Afridi et al., 2018).

[15]The correlation coefficient between 2014 $PM_{2.5}$ and the 2015-16 adult mortality measures calculated using the DHS-4 data is 0.0339 (SE: 0.0020, p-value: 0.00) for 15-49-year-old individuals and 0.4753 (SE: 0.0018, p-value: 0.00) for 50 years and older individuals at the district level.

rate since it reflects not just ambient air pollution but also other health-related district characteristics that proxy public health conditions like the number of hospitals and health policy interventions within the district.

We use the following district-specific adult mortality measure for each of the two consecutive age groups of 15-49 and 50 years or older to control for health-related time-varying district characteristics that are likely to explain the probability of child i's mortality incidence:

$$\textit{Adult mortality rate}^{\tau}_{dst} = \frac{\textit{Number of adults died}^{\tau}_{dst}}{\textit{Number of adults died}^{\tau}_{dst} + \textit{Number of adults alive}_{dst}}, \tag{2}$$

where *Number of adults died*$^{\tau}_{dst}$ is the total number of adults from survey households who died when being in the specified age-range over the past $\tau \in \{0, \ldots, 4\}$[16] years before the survey year t in district d of state s, and *Number of adults alive*$_{dst}$ is the total number of adults from surveyed households alive by the time of survey year t.[17] Since our qualitative results stay the same for different τ suggesting that the recall bias is evidently not present in our sample, we report the results for $\tau = 4$, i.e., adults who died in a district over the preceding four years before an interview.

The error term, ε_{ihdst}, captures the remaining unobserved, time-varying, and child-specific factors. We include an interaction between state and time fixed effects, η_{st}, to control for possible unobserved state-specific characteristics that vary over time.[18] The interaction term nonparametrically controls for state trends in under-five mortality, which is important in light of the time patterns observed in Figure 1. The under-five mortality rate remained stable at around 3% throughout the period from 1992-1993 to 2015-2016 for children in households using clean fuels, while their counterparts in households where polluting fuels are used face an almost quadrupled (in 1992-93) to doubled (in 2015-16) risk of mortality. Most of these results for

[16]The DHS survey records deaths over the past 4 years before the survey year at furthest and hence $\tau \in \{0, \ldots, 4\}$. Also taking advantage of data on causes of death recorded in the DHS, we calculate the mortality shares for those who died due to causes other than accident, violence, poisoning, and homicide or suicide, assuming that these five causes of death are less likely to be correlated with ambient air pollution or health system.

[17]We also used *Number of adults alive*$_{ds,t-\tau}$ as an alternative where $\tau \in \{0, \ldots, 4\}$, and the measure did not change significantly. The correlation of this alternative measure with the corresponding mortality measure using *Number of adults alive*$_{dst}$ is more than 0.98 for all five values of τ.

[18]Controlling for state $\times$ 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. For example, the government of India initiated the National Programme on Improved Chulha (NPIC) in the early 1980s to provide efficient cooking stoves to rural areas in an attempt to limit air pollution. NPIC became a nationwide program in the mid-1980s, and approximately 35 million improved cookstoves were distributed by the late 2000s. In late 2009, the Indian government also implemented a state-level voucher program, the National Biomass Cookstoves Initiative (NBCI). The pilot project distributed 12,000 improved cookstoves to households in the states of Jammu and Kashmir, Uttar Pradesh, Bihar, Madhya Pradesh, Jharkhand, Chhattisgarh, Karnataka, and Odisha (`https://pib.gov.in/newsite/printrelease.aspx?relid=94877`). Although existing studies such as Hanna et al. (2016) and Khandelwal et al. (2017) suggest the non-use of improved cookstoves distributed by the government programs in India, the set of state dummies in the baseline regressions addresses these and any other state-level policies.

under-five mortality are driven by infants. Across the different under-five age groups, mortality rates vary, and noticeable downward time trends apply to all age groups. The most at-risk group is neonates, and the associated neonatal mortality rate is the highest at 4.4% (1992-93) to 3.1% (2015-16). Post-neonatal mortality rate is next, at 3.0% (1992-93) to 1.3% (2015-16). Child mortality ranged from 1.1% (1992-93) to 0.3% (2015-16). It is worth noting that we cannot control for district or district $\times$ time fixed effects because one of our instruments, change in forest cover, is defined at the district level and only observed once across time. Heteroskedasticity-robust standard errors are clustered at the district level.

3 Identification

The key identification challenge is the potential endogeneity resulting from non-random use of polluting fuels. The empirical literature on the nexus between air pollution and health outcomes assumes the relationship to be unidirectional: that air pollution affects mortality and other human health outcomes, but not vice versa. For example, Duflo et al. (2008) document the potential impact of IAP on health, productivity, and ultimately long-term earnings. In our context, IAP and choice of fuel types for cooking can be affected by mortality, morbidity, and other health outcomes in addition to income. Given that low-income households can only afford cheaper fuel options which are frequently more polluting and have negative health and earnings implications, one valid concern is the simultaneity between the choice of cooking fuel and infant health outcomes in equation (1). Additionally, given that we use repeated cross-sectional data, we cannot control for fixed effects at the granular (or micro) level, e.g., child fixed effects or household fixed effects. Although we include a set of controls as detailed as we can in our setting, there could be still some unobserved characteristics that our controls fail to capture which can lead to an identification threat due to omitted variables bias. We address these concerns via the use of two instrumental variables (IVs): (i) the district forest cover as a region-specific characteristic and (ii) household ownership of agricultural land as a household-specific characteristic to generate plausible exogenous variation in the household fuel choice. These are discussed in detail below.

Forest cover. In the absence of data on prices of firewood and LPG, the main fuels for cooking in India, forest cover provides a proxy for the relative price (cost) of firewood at the district and/or village level.[19]

As we illustrate below, the forest cover is indeed relevant to generating meaningful variation in

[19] Kuo and Azam (2019) analyzes 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 the India Human Development Survey. Cooking fuel choice is shown to depend on different factors based on the geographical location of the household. While paved roads and peer effects significantly increase the probability of clean fuel choice for rural households, the bargaining power or economic status of women (captured via education, financial independence, and freedom to make decisions in general) and the price of LPG are critical determinants of clean fuel adoption in urban areas.

the opportunity cost of cooking fuel choice. Wood is the most widely used cooking fuel in India. Figure 2 shows that one-half of the Indian households covered by the four rounds of the DHS rely on wood as fuel for cooking. The second dominant cooking fuel is liquid petroleum gas (LPG) and natural gas, with a share of 32.4%. The other clean fuels account for only 1.4% (electricity is 0.9%, and biogas is 0.4%). Overall, based on our classification of cooking fuels, one-third of Indian households have been consuming clean fuels. The remaining two-thirds relied on polluting fuels for cooking over the past 25 years. The second and third most widely used polluting fuels are animal dung and crops, respectively, which are more prevalent among households with agricultural land. Forest cover generates variations in access to or availability of firewood for cooking, and households living in villages with forest cover use firewood twice as much as households in villages without (Pinto et al., 1985). Figure 3 illustrates India's district-wise forest cover in 2013 based on satellite data from the Planning Commission of India. While areas with dense forests are geographically concentrated in the northern, eastern, and southern regions, moderately forested areas are located in the central and central-east regions. From DHS-4 (2015-16), the share of households using solid fuels for cooking in the five states with forest cover above the country average in 2013 (90% in Jharkhand, 89% in Odisha, 87% in Assam, 85% in Chhattisgarh, and 81% in Madhya Pradesh) is substantially larger than the country average at 75%, suggesting that cooking fuel choice is indeed correlated with the location of forests. Figure 4a shows the positive relationship between forest cover and the use of polluting fuels for cooking in these five states at the district level.

Coincidentally, the under-five mortality rates in these same five states (6.0% in Jharkhand, 5.8% in Odisha, 6.2% in Assam, 6.0% in Chhattisgarh, and 7.0% in Madhya Pradesh) are also persistently higher than the country average of 5.4%. Figures 4b and 4c illustrate that the under-five mortality rate is positively associated with the use of polluting fuels for cooking and forest cover, respectively, for these five states. This suggests that forest cover is likely to be linked to infant mortality via biomass fuel use. Forest cover is a geospatial variable, and hence is likely to induce plausible exogenous variation in cooking fuel choice that is not correlated with the unobserved, time-varying, and child-specific shocks.

It is worth pointing out the reason as to why the forest cover at the *district* rather than at the village level is used as an instrument. To protect the identity of respondents, the DHS randomly displaces primary sampling units (PSUs, or villages/city blocks). This puts limits on the accuracy with which PSUs can be matched with Census locations and other datasets at sub-district (or *tehsil*) and village levels.[20] Figure B.2 illustrates the displacement strategy of PSU points in DHS-4, and how the displacement buffers may mask DHS survey respondents' residence location

[20] According to the description of the DHS GPS data provided by the DHS Program, the displacement is restricted so that the PSU points stay within the country, the DHS 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.

at the sub-district and village levels. In particular, the PSU point provided in the DHS-4, illustrated by the star, could be anywhere within the buffer that overlays multiple villages and even sub-districts. So the PSU point placed in a particular village could be, in fact, in a different village of the neighboring sub-district. Although the PSU point displacement is random, it can affect our empirical analysis since we match DHS data with satellite and Census data by location.

We obtained district-level satellite data on forest cover from the Planning Commission of India (replaced in 2015 by the National Institution for Transforming India–NITI Aayog) for three years–2007, 2011, and 2013.[21] The baseline regressions use the forest cover for 2013, which is the most recent and the closest period to the survey year used in the IV estimation. We also use forest cover data from 2011 as a robustness check. The forest cover is defined by forested area as a percentage of total geographical area, based on data from the NITI Aayog.[22]

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 forested area (hectare/person)[23] and percentage of the total geographical area of the village occupied by forests. 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 populations and inaccessible hilly geographic areas present a problem in the 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 0.68 in 2011 which shows that the Indian census- and satellite-based tree cover data are indeed different but quite comparable.

Agricultural land ownership. The second instrument we use for cooking fuel choice is household agricultural landownership - a binary variable that takes the value of 1 if the household owns land for agricultural purposes in a given year. Agricultural landownership generates exogenous variations in the opportunity cost of cooking fuel choice, as agricultural households are likely to consume their own agricultural crop waste and animal dung as cooking fuel which are classified as polluting.[24] A common argument is that infant mortality is negatively associated with

[21]The satellite-based forest cover data is collected every two years, and data on change in forested area from the previous round is included in each round of the satellite data. Thus, we calculate the forest cover for the years 2005 and 2009 using data for 2007 and 2011 data, respectively. This gives us a total of 5 years of satellite-based forest cover from 2005 to 2013 at the district level.

[22]In the Planning Commission dataset, forest cover refers to all land masses at 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 data on tree cover has been classified into four categories based on tree canopy density, including very dense forest, moderately dense forest, open forest, and scrub. We consider the first three of these forest types in our analysis excluding scrub.

[23]We find that the results are independent of whether we use total population or total area for normalizing forest cover, and we do not report results using per-capita forest cover because the results are essentially the same as those using forest cover as a percentage of total geographical area.

[24]Relatedly, crop residue contribute significantly to household energy needs for smallholder households in

ownership of agricultural land through a wealth/income channel given that agricultural production is a prominent source of household income and wealth. Although Figure 5a illustrates a negative relationship between infant mortality and household wealth, Figure 5b in contrast shows that agricultural land ownership status is non-monotonically associated with household wealth. The correlation between agricultural land ownership status and household wealth index is -0.07 (SE: 0.0020, p-value: 0.00). It suggests that agricultural landownership is not a good proxy for household wealth in our context, given numerous households that own small-sized and subsistence farms. This result is independent of how many groups we split households into, e.g., five groups or three groups.[25]

Validity of instruments. Four assumptions must be satisfied for us to interpret our results as causal. In Section 4.2, we formally test the validity of our instruments in terms of the first two assumptions: relevance (instruments are correlated with the endogenous regressor) and independence (instruments are uncorrelated with any confounders of exposure-outcome relationship). First, we adopt the Montiel-Olea and Pflueger weak IV test for overidentified models to assess the relevance of the two instruments in our model. This method yields the same robust F-statistic as the traditional tests developed by Staiger and Stock (1997), Stock and Yogo (2005), and Kleibergen and Paap (2006) in the just identified case (Olea and Pflueger, 2013). Comparing the effective F-statistic from the Montiel Olea-Pflueger weak IV test with the popular rule-of-thumb threshold of 10 can also be informative. We show that the endogenous regressor and the instruments are strongly correlated, suggesting that the relevance assumption is valid.

Second, in the light of our overidentified model, we test the restrictions that all IVs are uncorrelated with ε_{ihdst} and that the instruments satisfy the independence or orthogonality condition (Sargan, 1958; Hansen, 1982; Altonji et al., 2005). Third, although we formally test the relevance and independence of our instruments as discussed above, the exclusion restriction (instruments affect the outcome only through endogenous regressors) is not directly testable. As a result, in Section 4.3, we present additional evidence that the two instruments affect under-five mortality mainly through cooking fuel choice in our context.

Fourth, since we combine multiple instruments for a single endogenous treatment variable in a two-stage least squares (2SLS) setting, we need to verify that the well-known monotonicity assumption – the 2SLS estimate is a positively weighted average of local average treatment effects (LATEs) (Imbens and Angrist, 1994) – is satisfied. In our setting, the endogenous variable is polluting cooking fuel, which we instrument for using agricultural land ownership and forest cover. In this context, Mogstad et al. (2021) show that the 2SLS estimates can be a positively weighted average of LATEs under a "partial monotonicity" assumption.

developing countries such as Sub-Saharan Africa that rely on solid biomass fuels for cooking (Berazneva et al., 2018).

[25] While the DHS data does not directly include actual earnings, we find (unreported here) a negative and statistically significant relationship between agricultural land ownership and principal components of household wealth index which includes ownership of refrigerator, television, washing machine, electric fan, air conditioner or cooler and computer, after controlling for time and spatial fixed effects.

We verify the partial monotonicity condition in Table 2. Panel A of Table 2 reports coefficients from regressing the endogenous variable, polluting fuel use, on each instrument separately, along with a coefficient from regressing agricultural land ownership on forest cover. These regressions also control for baseline covariates. Column 1 shows that controlling for the covariates (but not the other instrument), the correlation between each instrument and the treatment is positive and statistically significant. It follows that the weights for each complier group must be consistent with the partial monotonicity assumption. Column 2 demonstrates that the partial correlation between the two instruments is essentially zero. This independence of instruments from each other also guarantees that the 2SLS weights are positive even if the traditional monotonicity assumption is violated. The joint distribution of the given two instruments, therefore, is sufficient to yield positive weights. Panel B of Table 2 presents the same results when forest cover is defined as a binary variable indicating whether the district's forest cover is above the mean. We do so to strictly follow the formal statistical tests proposed by Mogstad et al. (2021) for binary treatment and binary instruments. The results similarly suggest that the partial monotonicity assumption is satisfied when we have a binary measure of forest cover. Consistent with the strong positive correlations, the null hypothesis of negative weights is strongly rejected ($p = 0.000$), and the null hypothesis of positive weights is not rejected ($p = 1.000$). These findings provide credence to the validity of the partial monotonicity assumption in our context and allow us to interpret our IV/2SLS estimates as causal.

Table 3 presents summary statistics for cooking fuels, infant mortality, and other demographics. A majority (75%) of the households use polluting fuels, while the remaining households use clean fuels (electricity, LPG, natural, and bio-gas) fuels for cooking. Infant mortality rates are systematically higher than the national under-five mortality rate of 5.4%. Further, three-fourths of the children belong to households located in rural areas. Overall, across rural and urban areas, 69.8% of the mothers with children aged under five are in the 20-29 years old age bracket. Other socio-economic indicators exhibit an even distribution including household wealth, mother's education, gender of child, breastfeeding status, location where food is cooked, and type of house. The average household size is seven, and the average adult mortality rate is, respectively, 1.0% and 8.5% for those with ages 15-49 years and 50 years or older.

Table 4 provides the mean and standard deviation for the five outcome variables (infant mortality for different age groups) and the key explanatory variable (type of primary cooking fuel) by geographic region, along with the associated number of observations. Evidently, infant mortality rates and fuel choices vary significantly across regions nationwide. Out of ten states/union territories (UTs) with the highest incidence of under-five mortality and the highest share of polluting fuel-using households, six of them are common, including Madhya Pradesh, Assam, Bihar, Chhattisgarh, Odisha, and Jharkhand.

The number of observations in Tables 3 and 4 varies across variables mainly because some variables are not collected in different rounds of the DHS. For example, as we discussed earlier, the variable indicating where food is cooked was only collected in the DHS-4 survey. This leads to a substantial difference in the number of observations across specifications with and without this control. Other minor differences in observations are due to a few missing observations in the data, for example, variables on cooking fuel choice, type of house, mother's education level, and other variables used in measuring adult mortality rate. To examine whether our descriptive results in this section are affected by the samples, we summarize the data on a common sample and show that the results are not driven by the choice of specific sample (see Tables B.2 and B.3). It is worth noting that our main results from IV estimation are not affected by these different samples because IV estimates are based on the DHS-4 survey. In the next section, we discuss and investigate the sample differences for our non-IV or probit results.

We use a survival analysis to provide descriptive evidence that a child's vulnerability to indoor air pollution varies by age. Figure 6a presents the likelihood of surviving for children at different ages from clean and polluting fuel-using families and shows that (i) the proportion of surviving children is lower in polluting fuel-using households, and (ii) the incremental change in the likelihood of survival decreases as child's age increases. The Kaplan-Meier survival estimates suggest that the likelihood of survival is lower for polluting fuel-using families than the clean fuel-using ones, and the gap between the two lines is widest for younger infants (Figure 6b).

4 Results

We first present the estimated average marginal effects[26] of cooking fuel choice on child mortality using a probit model followed by the linear IV estimates.

[26] Marginal effects are generally 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. To obtain 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 a *discrete change* in predicted probabilities 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

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

where $f(\cdot)$ is the probability distribution function for a standardized normal variable, $\boldsymbol{x}_i' = (x_{i1} \cdots x_{ik} \cdots x_{iK})_{(1\times K)}$ is a vector of explanatory variables, and N and K are the number of observations and the number of regressors, respectively. Intuitively, for example, the AME of fuel choice reports the percentage point change in the probability of under-five mortality with a change in *polluting fuel for cooking* from 0 to 1. The standard errors of the AMEs in this paper have been computed using the Delta method.

4.1 Probit estimates

Table 5 presents the results obtained by estimating equation (1) as a pooled probit model for child mortality within different age groups under three separate specifications wherein more controls are added successively. In panel A for under-five mortality, the average marginal effect (AME) of polluting fuel for cooking ranges from 2.3 to 0.8 percentage points. The basic model, shown in Column (1), includes year and state fixed effects[27] and is estimated using DHS-1, DHS-2, and DHS-4, while the models shown in Columns (2) and (3) are based exclusively on DHS-4 since information on the location of cooking is only available in this last round. Given that the calculated marginal effects of polluting fuel use are consistently greater than zero and statistically significant at the 1 percent level for each specification, it is clear that IAP is positively linked with the mortality risk amongst children aged under five in India. We consider the last regression as our primary specification for probit estimates because the inclusion of state $\times$ year dummies controls for unobserved time-varying spatial factors including state attributes (e.g., characteristics of state magistrate, presence of government-sponsored programs covering child health services in the state, 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 potentially affect both under-five mortality and fuel choice.

To examine the relationship between cooking fuel choice and infant mortality at a more disaggregated level, we partition under-five children into two mutually exclusive sub-groups: child (ages 1-5) and infant (until age one). Comparing Columns (3) across panels B and C reveals that infants are more sensitive to the use of polluting fuel for cooking as the estimated coefficient on the fuel choice variable is an order of magnitude larger in the infant mortality regression. To further investigate the relationship at an even more disaggregated level, we divide infants into two separate age groups: post-neonatal (greater than 28 days but less than 1 year) and neonatal (until 28 days after birth). Comparing the last column in panels D and E shows that the effect of polluting fuel choice among infants is driven by neonatal mortality, while the association between polluting fuel use and post-neonatal mortality is not statistically significant. This is not a surprising result given that neonates, with undeveloped immune systems, spend the most time with their mothers and are therefore at the highest risk of exposure to IAP associated with dirty cooking fuel choices.[28]

[27] We do not introduce district fixed effects in equation (1) because the instrument of forest cover in 2013 is at the district level.

[28] The number of observations in Table 5 drops substantially when we move from Column (1) to Columns (2) and (3) because these last two columns use only the DHS-4 sample. To examine whether the changes in estimates are because of the sample change or because of the inclusion of demographic characteristics, we estimated the same sets of specifications on a common sample for each age group. The results, shown in Table B.4, suggest that the changes in estimates are mainly due to the inclusion of additional covariates. Restricting the sample to a common sample reduces the estimates in Column (1) to some extent, but the qualitative results remain the same.

The estimated coefficients on other explanatory variables are reported in Tables B.5–B.9 for under-five, child, infant, post-neonatal, and neonatal mortality, respectively. The estimates on the relationship between other characteristics and mortality for children under five years of age are qualitatively consistent with the literature (Naz et al., 2016). Zeroing in on the average marginal effects of the explanatory variables for neonatal mortality we find that the mortality risk is the highest when mothers do not breastfeed. In addition, neonatal mortality is positively associated with teenage motherhood and negatively related to the mother's education. Neonatal 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. Controlling for these variables, cooking outside is essentially the same as cooking in the living room in terms of their association with mortality. We also find that under-five, infant, post-neonatal, and neonatal mortality are higher in districts with high adult mortality rates. Finally, specifications with *district* and *district* $\times$ *time* fixed effects yield qualitatively identical estimates.

We also estimate linear probability models instead of probit regressions. The OLS results (Tables B.10-B.14) are qualitatively similar to our baseline results from probit models. Particularly, the use of polluting fuel for cooking is positively associated with under-five mortality, and the positive association is concentrated among infants and neonates.

4.2 Linear IV estimates

We first present evidence on how forest cover and agricultural land ownership relate to households' choice of cooking fuel types. The relationships are estimated using a linear model, where the dependent variable is a binary indicator for whether fuel choice is dirty or clean. The correlation coefficients for forest cover and agricultural land ownership with polluting fuel use for cooking are 0.0149 (SE: 0.0022, p-value: 0.00) and 0.1795 (SE: 0.0022, p-value: 0.00), respectively.

Column (1) of Table 6 reports the first-stage results when the indicator variable for household's agricultural land ownership is used as an IV. This variable has a positive and statistically significant impact on cooking fuel choice via the likely use of their own agricultural crop waste and animal dung as cooking fuel. Columns (2)–(6) of Table 6 present the estimates from the IV (2SLS) regressions for the five different age groups. The coefficient estimates on polluting fuel for cooking for under-five, infant, and neonatal mortality are indeed positive and statistically significant, ranging from 0.036 to 0.047.

Table 7 presents our main results when forest cover and an indicator variable for household's agricultural land ownership are jointly included as IVs. Column (1) in the top panel reports the first-stage results. The effective F-statistic of Olea and Pflueger (2013) is above the critical value of 17 for a worst-case bias of 10 percent (at 5 percent statistical significance) and is well above the rule-of-thumb threshold of 10, indicating that the two IVs provide plausible variations in fuel choice that can be leveraged to identify a causal effect of fuel choice on infant mortality. Columns (2)–(6) in the top panel present estimates from the IV (2SLS) regressions for the five different age groups. Heteroskedasticity-robust standard errors are clustered at the district instead of the PSU level, given that we use district-level forest cover as one of the instruments. The Hansen test implies no rejection of the null hypothesis of valid instruments and suggests that the excluded IVs are exogenous. The coefficient estimates on polluting fuel for cooking for under-five, infant, and neonatal mortality are positive and statistically significant at the 5 percent level, ranging from 0.030 to 0.041. Specifically, the estimated effects on under-five, infant, and neonatal mortality rates are 0.040 (standard error = 0.020), 0.041 (standard error = 0.019), and 0.030 (standard error = 0.015), respectively. These suggest that approximately 27 under-five children and infants and about 20 neonates would have been saved per 1,000 live births economy-wide if all households used clean fuels for cooking.[29,30]

Our analysis points to a non-trivial impact of polluting fuel use on under-five, infant, and neonatal mortality in India, and we need to ensure that these results hold up both under further technical scrutiny as well as when accounting for a couple of unique socio-demographic features of Indian households. Technically, we need to confirm that the conventional heteroskedasticity-robust variance of the 2SLS estimator is not misleading since our local average treatment has heterogeneous effects. Specifically, the concern might be that the moment condition evaluated by the two-stage least squares (2SLS) estimand — a positively-weighted average of multiple local average treatment effects (LATEs) given more than one instrument — is misspecified (Lee, 2018). As a result, Table 7 also reports heteroskedasticity standard errors robust to multiple-LATEs which confirms that the statistical significance of our key explanatory variable remains unchanged.

In terms of the unique socio-demographic features, we first isolate households that exclusively rely on biomass fuels (which coincidentally also happens to be the cooking fuel choice of the poor in rural India), and second, differentiate by household size which can determine either the feasibility of switching to cleaner fuels or reducing the burden of cooking on mothers. We discuss these in detail below.

Singling out biomass fuels. Studies such as Imelda (2018, 2020) that causally estimate the impact of IAP on infant mortality focus exclusively on the impact of a switch of cooking fuel from

[29] The counterfactual mortality estimate per 1000 live births is derived based on the difference between (i) the predicted under-five mortality rate (= 0.040583) using observed controls including cooking fuel choice and the IV (2SLS) coefficient estimates, and (ii) the predicted under-five mortality rate (= 0.013678) using observed controls except for cooking fuel which we assume for this exercise that every household refrains from using polluting cooking fuel. We then rescale to give (40.583 – 13.678) 26.905 deaths per 1000 live births. Upon rounding, we arrived at the 27 deaths per 1000 live births.

[30] Table B.15 shows the full version of Table 7, providing coefficient estimates on all covariates.

kerosene to LPG, thereby excluding the dirtiest biomass fuels (including animal dung, agricultural waste, straw, shrubs, grass, and firewood) commonly used in the developing world. However, evidence suggests that biomass fuels, being the cheapest, are not only the preferred choice for a vast majority of poor households they are also far more dangerous to human health than kerosene (e.g., Fullerton et al., 2008). Hence, it is important to check for the risk of under-five, infant, and neonatal mortality due to the use of only biomass fuels, doubly so since these are the key fuel choices influenced by our IVs. The causal effects of biomass fuel for cooking on under-five, infant, and neonatal mortality rates range from 0.026 to 0.037 (Columns (2)–(6) in the bottom panel of Table 7). Unsurprisingly, these numbers are lower than their respective counterparts above (0.030-0.041) since biomass fuels are a subset of our complete list of polluting fuels.[31] The magnitudes of the causal effects of biomass fuels are nonetheless very close to the causal effect of the full list, suggesting the central importance of biomass fuel choice in determining the impact of IAP on child mortality.

Additional heterogeneity results by household size. The probit model suggests that the risk of under-five mortality, as well as those of the four age sub-groups, are inversely and significantly related to household size (Tables B.5–B.9). This is an intriguing relationship that requires further unpacking. We identify three channels via which the size of a household might guide either cooking fuel choice or the burden of cooking and childcare for resident mothers. These include (i) a smaller household may find it easier to switch to cleaner fuels - LPG as an example - since pre-fabricated gas stoves limit the size of cooking utensils that can be used (and therefore the amount of food that can be cooked in a given sitting), (ii) if a child's exposure to IAP is due primarily to time spent near cooking locations with the mother, larger household size may arguably relieve some of the burden of cooking or childcare on mothers and mitigate the child's mortality risk due to IAP, and (iii) if cooking outdoors is more practical for a larger household (due to a larger volume of cooking required), then a child's exposure to IAP will also decrease even if the time burden of cooking duties for the mother is not shared. Our data permits a direct check on the possibility (iii), and we find that household size is negatively correlated with cooking in a separate building (ρ: -0.003, SE: 0.002, p-value: 0.162) and outdoors (ρ: -0.005, SE: 0.002, p-value: 0.027).

Because of a potential non-linearity, we utilize data splits to analyze heterogeneous LATEs by household size. We focus on the sub-population of households with fewer than twelve members since such households constitute 96% of our sample. In our main analysis, we classify the sample of households with twelve or fewer members into three equally spaced groups (1-4, 5-8, and 9-12 members). The joint F-statistic on the excluded instruments indicates that the two IVs introduce meaningful variations in fuel choice in all regressions estimated on these sub-populations. We

[31] The reason that the dirtiest fuel—biomass fuel—has a mortality impact smaller than overall polluting fuels is that we include the other polluting fuels in the non-biomass category, including kerosene, coal or lignite, and charcoal, for this regression, which raises the adverse health impact of the reference category. In other words, we compare the health impact of the dirtiest fuel with that of other dirty fuels and identify their differential effect.

find that the effect of under-five, infant, and neonatal mortality rates is 0.059, 0.056, and 0.048, respectively, in households with 5-8 members (Columns (2), (4), and (6) in panel B of Table 8). Panels C and D of Table C.2 show that this result is driven by households with 5-6 members, which is smaller than the average household size of 6.9 members (see Table 3), as the coefficient estimates essentially becomes zero for households with 7-8 members. The coefficient estimates on IAP in under-five, infant, post-neonatal, and neonatal mortality for households with 1-4 members are positive and statistically significant at least at the 5% level. However, the Hansen test rejects the null hypothesis at the 5% and 10% levels, suggesting that the instruments are not valid for these households. This lends credence to our hypothesis (i) above that child mortality in the smallest household is not necessarily caused by polluting fuel use since these households can easily switch to pre-fabricated gas stoves that only accommodate smaller cooking utensils. The effect of IAP on infant mortality for all under-five age groups is essentially zero for larger households with 9-12 members. Results thus indicate that infants are subject to a greater risk of mortality due to IAP in households with 5-6 members - a household size that's large enough to preclude a switch to pre-fabricated gas stoves and small enough where mothers need to engage in cooking and childcare simultaneously.[32]

Additional heterogeneity by child's gender. We also analyze heterogeneity by child's gender, given a preference for sons in India (Arnold et al., 1998; Jayachandran and Pande, 2017; Jayachandran, 2023). There are two possibilities. Polluting fuels for cooking could have more adverse effects on boys' health if the only source of differential treatment comes from parents spending more time caring for sons than daughters, including time spent near cooking locations. Another possibility is that there are multiple sources of differential treatment by child's gender, including for example medical treatment, or pollution exposure adjustment if respiratory diseases arise. The mortality effect we observe sums up the combined outcomes of pollution exposure and ex-post medical treatment and/or exposure adjustments. As shown in Table 9, we find that the adverse mortality effect of polluting fuel use is fact concentrated on girls. The coefficient estimate on polluting fuels for cooking for mortality of under-five girls is 0.078, implying that 53 lives per 1,000 live births of girls are lost within five years of life due to the use of polluting cooking fuels. This result is consistent with descriptive results from India's DHS data, suggesting that under-five mortality is higher among girls than boys.[33] This may arise if (i) only girls are more exposed to polluting fuels, or (ii) both boys and girls are exposed but boys are more likely to receive ex-post care if the disease does arise, or both. This mechanism of girls being overlooked or receiving less care from parents could be more compelling as previous studies in India suggest that women,

[32] In the Appendix C, we check whether size classification affects our results of heterogeneous LATE by household size. The results with different size classifications consistently suggest that the adverse health impact of IAP on neonatal mortality exacerbates as the household becomes smaller, except for small families with up to 3 members for which the effects of IAP on neonatal mortality disappear. It is worth noting that the coefficient on IAP in the under-five, infant, and post-neonatal regressions for some households becomes statistically significant, but the exogeneity test fails for all those regressions of under-five mortality other than neonates.

[33] https://www.unicef.org/india/what-we-do/newborn-and-child-health

especially young girls, are neglected and mismanaged during illness more than boys. It is also consistent with the fact that one of the dominant causes of excess female deaths in India is respiratory diseases potentially due to polluting fuel use for cooking (Anderson and Ray, 2010).

Additional heterogeneity by status of male siblings. Given a preference towards sons (e.g., Jayachandran, 2023), one might consider that the effective care burden of a mother is more closely linked to the number of male siblings of a child rather than the size of the household. Thus, we estimate the mortality effects of polluting fuels with and without male siblings separately. As shown in Table 10, the qualitative results are generally the same for children with and without male siblings. The quantitative results suggest that the mortality effect is much more pronounced for children with male siblings, as the coefficient estimates on under-five and infant mortality in the top panel are almost twice as large as those in the bottom panel in magnitude. For neonatal mortality, the effect is around three times larger for neonates with male siblings than those without any male siblings.

4.3 Assessing the exclusion restriction

We assess the validity of the exclusion restriction for our instruments via several tests below.

Zero-first-stage test. First, an informal but informative test, the zero-first-stage test, can be used to lend some confidence that the exclusion restriction is reasonable (Van Kippersluis and Rietveld, 2018). The method was first used by Bound and Jaeger (2000) and later popularized by Altonji et al. (2005) and Angrist et al. (2010). The intuition of the approach is that the reduced-form effect of the instrument on the outcome should be zero in a subsample for which the first-stage effect of the instrument on the treatment variable is zero if the exclusion restriction is satisfied. Although this analysis cannot be used to verify the exclusion restriction, we use this to assess whether the exclusion restriction is plausible. We use a subsample of small households from urban areas as the zero-first-stage sample since these households have limited access to forests as a source of firewood and limited capability of growing crops and using agricultural crop residues for cooking. Notable here is our focus on small household size as an additional dimension that determines the relevance of our instruments to cooking fuel choice since small, i.e., labor-constrained households, are unlikely to engage in farming (even if they own land) and collect firewood (even if they reside close to a forest). Thus, we define the zero-first-stage subsample as households from urban areas with three members or less, e.g., a child and two parents.

Table 11 presents the results from this analysis. The first-stage effects of agricultural land ownership (Column (1)) and forest cover and agricultural land ownership (Column (2)) on polluting fuel use for cooking is essentially zero for small households from urban areas (panel (b)), confirming that they are a valid zero-first-stage group. As expected, the effect of the instrument(s) on cooking fuel choice is statistically significant, and instrument(s) are relevant (according to the Olea and Pflueger (2013) test for weak instruments) for both the full and remaining samples (panels (a) and (c)). Consistent with the validity of the exclusion restriction, the direct effects of agricultural land ownership (Column (1)) and forest cover and agricultural land ownership (Column (2)) on under-five mortality are statistically insignificant for small households from urban areas (panel (e)).

Additional indirect tests. Second, we estimate the association between determinants of child mortality, other than cooking fuel choice, through the following regressions:

$$y_{ihdst} = \alpha + \beta_1 \textit{Forest Cover}_{dst} + \beta_2 \textit{Agricultural Land Ownership}_{hdst} + \\ + \textit{Household}_{hdst}\boldsymbol{\gamma} + \textit{Mother}_{ihdst}\boldsymbol{\lambda} + \textit{Child}_{ihdst}\boldsymbol{\delta} + \textit{District}_{dst}\boldsymbol{\pi} + \eta_{st} + \varepsilon_{ihdst}, \tag{3}$$

where y_{ihdst} is a characteristic of child i reported in the DHS data including a child's breastfeeding status — one of the most important determinants of under-five mortality found in our baseline analysis. As a proxy for the mother's health-seeking behavior, we introduce the number of antenatal visits during pregnancy and the number of injections (of any kind) received in the last 12 months in this specification. All other variables, including our baseline instruments $\textit{Forest Cover}_{dst}$ and $\textit{Agricultural Land Ownership}_{hdst}$, remain identical to those in equation (1).

Table B.16 reports the results from estimating equation (3) using OLS. The results show that our baseline instruments are not correlated with variables commonly expected to be associated with child mortality other than cooking fuel choice since the Olea and Pflueger (2013) test values for the instruments imply that they are weak in those specifications. The coefficient estimate on forest cover is weakly significant at the 10% level in the maternal health regression. These findings lend credence to the exclusion restriction in our context, given that our instruments do not introduce noteworthy variation in the determinants of child mortality other than through the cooking fuel choice.

As a final test, we consider the possibility that our instruments may be relevant to explanatory variables of child mortality other than those investigated above and the cooking fuel choice. We examine this in Table B.17 by using household size and mother's education as examples of household and mother's characteristics, respectively, that are included in equation (1) as determinants of child mortality in addition to polluting fuel use for cooking. We focus on under-five mortality here since it is our most general child mortality outcome; however, the results remain the same for mortality outcomes for other child age groups. Our baseline analysis suggests that under-five mortality is associated with smaller households and less educated mothers (see Table B.10), and that our instruments are relevant to these two variables since the effective F-statistic on instruments are large enough (Columns (1) and (3) of Table B.17). However, IV estimation results suggest that household size and mother's educational attainment instrumented by agricultural land ownership status and forest cover do not affect under-five mortality, unlike cooking fuel choice that we find in our baseline analysis (Columns (2) and (4) of Table B.17). This lends credibility to the exclusion restriction of the instruments in our context.

5 Robustness checks

A battery of alternative specifications serves as robustness checks including (i) probit specifications with instruments, (ii) a more refined definition of polluting cooking fuels, (iii) alternative forest cover data, and (iv) an alternative definition of mortality outcomes for children in different age groups.

5.1 Probit specification with instruments

We re-estimate the causal effect of cooking fuel choice on infant mortality using (two-step) IV probit regressions as an alternative to the linear IV regression. We find that the IV probit provides the same qualitative conclusion as the 2SLS regression, verifying that the results are robust to an alternative approach. Table A.1 presents the parameter estimates derived from the IV probit regression for under-five, child, infant, post-neonatal, and neonatal mortality (with the same specification as used in Table 7 where both forest cover and agricultural land ownership are used as IVs). The coefficient estimates for the use of dirty cooking fuel on under-five, infant, and neonatal mortality are positive and statistically significant and reinforce the positive relationship between cooking fuel choice and the risk of mortality for the youngest children.

5.2 Redefining polluting cooking fuel

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.2 shows that the effects of a household switching to a fuel type that is dirtier by one level down the energy ladder or increasing the dirtiness level of cooking fuel by one unit on under-five, infant, and neonatal mortality rates are 0.007, 0.007, and 0.005, respectively. Notice that although the key regressor here (dirtiness level of cooking fuels) is a categorical variable, our results remain remarkably similar to the baseline results that use a binary variable that captures the use of polluting fuel for cooking.

5.3 Alternative forest cover data

We leverage satellite- and census-based data on forest cover (% of geographical area) in 2011 to test whether we can identify a positive impact of polluting fuel use on under-five, infant, and neonatal mortality incidences. Using data on 2011 satellite-based forest cover and tree cover from the 2011 Indian Census as alternates to a 2013 satellite-based forest cover, we find that our results remain 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.3 and A.6, respectively). An exception is the under-five regression, in which the coefficient estimate on polluting fuel use is no longer statistically significant when census-based forest is used. However, the impact of polluting cooking fuel choices on infant and neonatal mortality remains significant, indicating that the IAP impact on under-five mortality is mainly driven by its impact on neonates.[34]

5.4 Redefining the mortality outcomes

In our baseline analysis, we define the outcome of $\textit{Child Mortality}_{ihdst}$ for each of the five age groups as an indicator variable that takes a value of 1 if the mortality occurred over the given age period, and 0 otherwise. But, according to this definition, the mortality dummy for a particular age group includes deaths that happened in different age groups, even if those periods during which the deaths occurred are mutually exclusive. For example, in our baseline definition of neonatal mortality, we code children who died younger than 28 days as one and everyone else, such as those who died older than 28 days, as zero.[35] In this sub-section, we redefine our four indicator variables of mortality for each age group by removing children who died outside of the respective age group. First, in the alternative definition of neonatal mortality, we remove children who died older than 28 days and keep everything else the same as the baseline definition. Second, children who died younger than 27 days and between 1 and 5 years of age are dropped from the alternative post-neonatal mortality measure. Third, children who died older than a year are removed from the infant mortality measure. Fourth, we remove children who died younger than a year old in this alternative definition of our child mortality measure. The baseline definition of under-five mortality remains the same as it covers all five years preceding the survey - in effect, the under-five mortality measure is indicated by a value of one if the child died within the first five years of life and zero otherwise.

Table A.5 presents the probit estimates of the redefined mortality outcomes. Given that we drop deaths that occurred in other age groups from each age-specific mortality variable, the number of observations is slightly reduced compared to the baseline model in Table 5. However, the sign and the magnitude of the average marginal effects (AMEs) and statistical significance of estimates remain robust. Note that since we did not redefine the under-five mortality indicator, we do not check for the robustness of the under-five mortality regression. As before, our results show that polluting fuel use for cooking is positively associated with child mortality, and it is concentrated amongst neonates. Interestingly, the positive relationship for post-neonates was not

[34] Note that the Hansen J statistic suffers slightly when we use census-based forest cover, which could be due to issues with the forest cover data from the Census as outlined in Foster and Rosenzweig (2003).

[35] Similarly, for post-neonatal mortality, we code children who died between 28 days and a year old as one and everyone else, such as those who died younger than 27 days and between 1 and 5 years of age, as zero. For infant mortality, we code children who died at an age younger than a year old as one and everyone else, such as those who died older than a year, as zero. For child mortality, we code children who died between 1 and 5 years of age as one and everyone else, such as those who died younger than a year old, as zero.

statistically significant in the baseline analysis but is now statistically significant at the 10% level.

Table A.6 reports our main IV results, showing that our results on the effects of polluting fuel use on age-specific mortality are highly robust to the alternative definition of mortality outcomes. We did not report the first-stage results and results from under-five mortality because they are similar to the baseline. In addition to the estimation results, the test statistics also qualitatively remain the same.

5.5 Controlling for breastfeeding duration

We control for the child's breastfeeding status in our baseline regressions. However, previous evidence suggests that girls in India tend to be breastfed for a shorter period than boys (Jayachandran and Kuziemko, 2011; Dutta et al., 2022). As a result, breastfeeding status in itself might not capture this gender difference in breastfeeding *duration*. While the DHS data reports the duration over which a child was breastfed, the information is missing for about one-fourth of the children for whom breastfeeding status is available. This is the main reason behind our focus on the child's breastfeeding status in our baseline analysis. However, we check the robustness of our results by controlling for the child's breastfeeding duration instead of the child's breastfeeding status, focusing on our main IV results. As shown in Table A.7, the qualitative results are strongly robust.

6 Conclusion

Many countries, especially in the developing world, have attempted to resolve IAP by implementing programs to switch to clean fuels and adopt improved cookstoves. Notable amongst these are (i) shifting fuel subsidies like The Mega Conversion Program from Kerosene to LPG in Indonesia and the Promotion of Electric Induction Stoves to Reduce LPG Consumption in Ecuador, (ii) subsidies for improved technologies like The Fondo de Inclusión Social Energético LPG Subsidization Program in Peru, LPG subsidies through the Ujjwala Program in India and the Rural LPG Program in Ghana, (iii) Microfinance for LPG Equipment in Kenya, (iv) consumer credit initiatives like the Sales Offers for Improved Cookstoves in Uganda and Pay-as-you-Go LPG in Kenya, and (v) conditional cash transfers for LPG use in India. The existing studies focus on investigating the successes and failures of such programs in specific contexts and suggest that the health benefits are limited to the short term. Countries fail to sustain the energy transition in the long run due to a variety of factors, such as the absence of internationally-recognized clean cookstove standards and limited in-country testing capabilities, lack of awareness among households regarding the benefits of clean cookstoves and fuels, issues related to distribution and supply chain, especially in rural areas and finally, the high initial cost of adopting clean cookstoves (Cordes, 2011).

Due to this, almost half of the world's population in developing countries continues to depend on dirty cooking fuels, which constitute the largest source of poor indoor air quality. The consequence of dirty fuel choice, primarily biomass, on respiratory health is significant, with the youngest household members - children under five - being the most vulnerable. Yet, no causal estimate exists on the true impact of dirty fuel choice on under-five and infant mortality, even though diseases attributable to indoor air pollution account for more deaths than malaria and tuberculosis combined globally. Leveraging unique and large-scale household survey data from 1992 to 2016 and geospatial information on forest cover in India, where both the use of dirty cooking fuel and child mortality is high, we find that the use of polluting fuels for cooking or IAP has a significant and robust impact on under-five and infant mortality mainly via its effect on neo-neonatal mortality.

Aside from being the first causal study to employ large-scale nationally representative demographic survey data to identify the use of solid fuels as a contributor to under-five mortality, we uncover three critical relationships. First, we show that the effects of polluting fuel use for cooking on neonatal, infant, and under-five mortality rates are 0.030, 0.041, and 0.040, respectively. Second, neonatal mortality is the highest in relatively small households with 5-6 members, where mothers share the dual burden of child-rearing and cooking. Third, our finding suggests that it is mainly biomass fuels that drive the adverse impact of polluting fuel use on child mortality in our context. This paper is the first to document that the effect of IAP on under-five mortality is significantly heterogeneous by both a child's age and gender and household size.

We conclude with some caveats and directions for future research. First, our analysis of the causal impact of IAP on under-five mortality is based on an indirect indicator of IAP, i.e., the type of primary cooking fuel. Using direct measures of IAP (CO and PM concentrations in homes) recorded by 24-hour carbon monoxide readings might provide more accurate estimates due to less measurement error but data of this sort is generally unavailable in the developing world. Second, while we focus on the causal impact of IAP on infant mortality, IAP has adverse impacts on other socio-economic and health outcomes. Future research on the impact of cooking fuel choice on the productivity of children and adults - proxied by school attendance and labor market participation respectively, can uncover other dimensions of the economic cost of using dirty cooking fuels.

References

Afridi, Farzana, Taryn Dinkelman, and Kanika Mahajan. 2018. "Why are fewer married women joining the work force in rural India? A decomposition analysis over two decades." *Journal of Population Economics*, 31: 783–818, DOI: http://dx.doi.org/10.1007/s00148-017-0671-y.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "An evaluation of instrumental variable strategies for estimating the effects of Catholic schooling." *Journal of Human Resources*, 40(4): 791–821, URL: http://www.jstor.org/stable/4129541.

Anderson, Siwan, and Debraj Ray. 2010. "Missing women: Age and disease." *The Review of Economic Studies*, 77(4): 1262–1300, DOI: http://dx.doi.org/10.1111/j.1467-937X.2010.00609.x.

Angrist, Joshua, Victor Lavy, and Analia Schlosser. 2010. "Multiple experiments for the causal link between the quantity and quality of children." *Journal of Labor Economics*, 28(4): 773–824, DOI: http://dx.doi.org/10.1086/653830.

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, DOI: http://dx.doi.org/10.1111/ecoj.12273.

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, DOI: http://dx.doi.org/10.1542/peds.108.4.e67.

Arnold, Fred, Minja Kim Choe, and Tarun K. Roy. 1998. "Son preference, the family-building process and child mortality in India." *Population Studies*, 52(3): 301–315, DOI: http://dx.doi.org/10.1080/0032472031000150486.

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, DOI: http://dx.doi.org/10.1086/684582.

Barron, Manuel, and Maximo Torero. 2017. "Household electrification and indoor air pollution." *Journal of Environmental Economics and Management*, 86: 81–92, DOI: http://dx.doi.org/10.1016/j.jeem.2017.07.007.

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, DOI: http://dx.doi.org/10.1186/1471-2458-10-491.

Beach, Brian, and W. Walker Hanlon. 2018. "Coal smoke and mortality in an early industrial economy." *The Economic Journal*, 128(615): 2652–2675, DOI: http://dx.doi.org/10.1111/ecoj.12522.

Benshaul-Tolonen, Anja. 2019. "Local industrial shocks and infant mortality." *The Economic Journal*, 129(620): 1561–1592, DOI: http://dx.doi.org/10.1111/ecoj.12625.

Berazneva, Julia, David R. Lee, Frank Place, and George Jakubson. 2018. "Allocation and valuation of smallholder maize residues in Western Kenya." *Ecological Economics*, 152: 172–182, DOI: http://dx.doi.org/10.1016/j.ecolecon.2018.05.024.

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, DOI: http://dx.doi.org/10.1016/S0140-6736(03)13779-8.

Bound, John, and David A. Jaeger. 2000. "Do compulsory school attendance laws alone explain the association between quarter of birth and earnings?" In *Research in Labor Economics*. Vol. 19: Bingley, West-Yorkshire: Emerald Group Publishing Limited, 83–108, DOI: http://dx.doi.org/10.1016/S0147-9121(00)19005-3.

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, URL: http://stacks.cdc.gov/view/cdc/6914/cdc_6914_DS1.pdf.

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, DOI: http://dx.doi.org/10.1111/ecoj.12285.

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, DOI: http://dx.doi.org/10.1257/pol.20150085.

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, DOI: http://dx.doi.org/10.1257/app.20160436.

Chen, Yuyu, Avraham Ebenstein, Michael Greenstone, and Hongbin Li. 2013. "Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy." *Proceedings of the National Academy of Sciences*, 110(32): 12936–12941, DOI: http://dx.doi.org/10.1073/pnas.1300018110.

Cordes, Leslie. 2011. *Igniting change: A strategy for universal adoption of clean cookstoves and fuels*. Global Alliance for Clean Cookstoves, URL: http://cleancooking.org/reports-and-tools/igniting-change-a-strategy-for-universal-adoption-of-clean-cookstoves-and-fuels/.

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, DOI: http://dx.doi.org/10.1093/oxfordjournals.aje.a009540.

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, DOI: http://dx.doi.org/10.1257/aer.20131002.

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, DOI: http://dx.doi.org/10.1136/jech.2006.043133.

Duflo, Esther, Michael Greenstone, and Rema Hanna. 2008. "Indoor air pollution, health and economic well-being." *SAPI EN. S. Surveys and Perspectives Integrating Environment and Society*, 1(1), URL: http://journals.openedition.org/sapiens/130.

Dutta, Swati, Sunil Kumar Mishra, and Aasha Kapur Mehta. 2022. "Gender discrimination in infant and young child feeding practices in India: Evidence from NFHS-4." *Indian Journal of Human Development*, 16(2): 286–304, DOI: http://dx.doi.org/10.1177/09737030221117867.

Ebenstein, Avraham, Maoyong Fan, Michael Greenstone, Guojun He, and Maigeng Zhou. 2017. "New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy." *Proceedings of the National Academy of Sciences*, 114(39): 10384–10389, DOI: http://dx.doi.org/10.1073/pnas.1616784114.

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, DOI: http://dx.doi.org/10.1017/S1355770X12000010.

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, DOI: http://dx.doi.org/10.1186/1476-069X-13-113.

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, DOI: http://dx.doi.org/10.1353/dem.2001.0003.

Foster, Andrew D., and Mark R. Rosenzweig. 2003. "Economic growth and the rise of forests." *The Quarterly Journal of Economics*, 118(2): 601–637, DOI: http://dx.doi.org/10.1162/003355303321675464.

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, DOI: http://dx.doi.org/10.1016/j.trstmh.2008.05.028.

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, DOI: http://dx.doi.org/10.2307/1924934.

Ghosh, Pronay, Vasundhara Chatterjee, Avike Paul, Debarati Ghosh, and Zakir Husain. 2024. "Reducing energy poverty: How to empower women and switch to clean fuel in India?" *Energy Research & Social Science*, 110: 103444, DOI: http://dx.doi.org/10.1016/j.erss.2024.103444.

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, DOI: http://dx.doi.org/10.1016/j.jeem.2009.03.001.

Graff Zivin, Joshua, and Matthew Neidell. 2012. "The impact of pollution on worker productivity." *American Economic Review*, 102(7): 3652–3673, DOI: http://dx.doi.org/10.1257/aer.102.7.3652.

Graff Zivin, Joshua, and Matthew Neidell. 2018. "Air pollution's hidden impacts." *Science*, 359(6371): 39–40, DOI: http://dx.doi.org/10.1126/science.aap7711.

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, URL: http://www.bjcancer.org/Sites/Uploaded/File/2016/11/30636160943288567782796154 2.pdf.

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, DOI: http://dx.doi.org/10.1257/pol.20140008.

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, DOI: http://dx.doi.org/10.1016/j.jpubeco.2014.10.004.

Hansen, Lars Peter. 1982. "Large sample properties of generalized method of moments estimators." *Econometrica*, 50(4): 1029–1054, DOI: http://dx.doi.org/10.2307/1912775.

Heltberg, Rasmus. 2005. "Factors determining household fuel choice in Guatemala." *Environment and Development Economics*, 10(3): 337–361, DOI: http://dx.doi.org/10.1017/S1355770X04001858.

Heltberg, Rasmus, Thomas Channing Arndt, and Nagothu Udaya Sekhar. 2000. "Fuelwood consumption and forest degradation: A household model for domestic energy substitution in rural India." *Land Economics,* 213–232, DOI: http://dx.doi.org/10.2307/3147225.

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, URL: http://www.undp.org/sites/g/files/zskgke326/files/publications/chapter3.pdf.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and estimation of local average treatment effects." *Econometrica*, 62(2): 467–475, DOI: http://dx.doi.org/10.2307/2951620.

Imelda. 2018. "Indoor air pollution and infant mortality: A new approach." *AEA Papers and Proceedings*, 108: 416–421, DOI: http://dx.doi.org/10.1257/pandp.20181119.

Imelda. 2020. "Cooking that kills: Cleaner energy access, indoor air pollution, and health." *Journal of Development Economics*, 147: 102548, DOI: http://dx.doi.org/10.1016/j.jdeveco.2020.102548.

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, DOI: http://dx.doi.org/10.1086/705554.

Jayachandran, Seema. 2009. "Air quality and early-life mortality evidence from Indonesia's wildfires." *Journal of Human Resources*, 44(4): 916–954, DOI: http://dx.doi.org/10.

3368/jhr.44.4.916.

Jayachandran, Seema. 2023. "Ten facts about son preference in India." NBER Working Paper No. 31883, DOI: http://dx.doi.org/10.3386/w31883.

Jayachandran, Seema, and Ilyana Kuziemko. 2011. "Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India." *The Quarterly Journal of Economics*, 126(3): 1485–1538, DOI: http://dx.doi.org/10.1093/qje/qjr029.

Jayachandran, Seema, and Rohini Pande. 2017. "Why are Indian children so short? The role of birth order and son preference." *American Economic Review*, 107(9): 2600–2629, DOI: http://dx.doi.org/10.1257/aer.20151282.

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, DOI: http://dx.doi.org/10.1093/ej/uez021.

Khandelwal, Meena, Matthew E. Hill Jr., Paul Greenough, Jerry Anthony, Misha Quill, Marc Linderman, and H.S. Udaykumar. 2017. "Why have improved cook-stove initiatives in India failed?" *World Development*, 92: 13–27, DOI: http://dx.doi.org/10.1016/j.worlddev.2016.11.006.

Kleibergen, Frank, and Richard Paap. 2006. "Generalized reduced rank tests using the singular value decomposition." *Journal of Econometrics*, 133(1): 97–126, DOI: http://dx.doi.org/10.1016/j.jeconom.2005.02.011.

Kuo, Ying-Min, and Mehtabul Azam. 2019. "Household cooking fuel choice in India, 2004-2012: A panel multinomial analysis." IZA Discussion Paper No. 12682, DOI: http://dx.doi.org/10.2139/ssrn.3468630.

Lee, Seojeong. 2018. "A consistent variance estimator for 2SLS when instruments identify different LATEs." *Journal of Business & Economic Statistics*, 36(3): 400–410, DOI: http://dx.doi.org/10.1080/07350015.2016.1186555.

Lee, YuJung Julia, Zakir Husain, and Mousumi Dutta. 2023. "Does improved cooking fuel empower women? Evidence from India." *Sustainable Development*, DOI: http://dx.doi.org/10.1002/sd.2695.

Liu, Ziming, Jia Li, Jens Rommel, and Shuyi Feng. 2020. "Health impacts of cooking fuel choice in rural China." *Energy Economics*, 89: 104811, DOI: http://dx.doi.org/10.1016/j.eneco.2020.104811.

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, DOI: http://dx.doi.org/10.1016/j.reseneeco.2006.02.002.

Mogstad, Magne, Alexander Torgovitsky, and Christopher R. Walters. 2021. "The causal interpretation of two-stage least squares with multiple instrumental variables." *American Economic Review*, 111(11): 3663–3698, DOI: http://dx.doi.org/10.1257/aer.20190221.

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, DOI: http://dx.doi.org/10.3368/jhr.46.1.154.

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, DOI: http://dx.doi.org/10.1186/s12940-016-0138-8.

Olea, José Luis Montiel, and Carolin Pflueger. 2013. “A robust test for weak instruments.” *Journal of Business & Economic Statistics*, 31(3): 358–369, DOI: http://dx.doi.org/10.1080/00401706.2013.806694.

Pinto, A. G., P. Khanna, and C. N. Pandey. 1985. “Gujarat wood balance study 1984.” Unpublished, 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*. Vol. 202, Stanford: Stanford Institute for Theoretical Economics, URL: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=a15741bf804c746b7bb06ae12ee04af7e355e786.

Rehfuess, Eva A., N. G. Bruce, and K. R. Smith. 2011. “Solid fuel use: Health effect.” *Encyclopedia of Environmental Health*, 5: 150–161, DOI: http://dx.doi.org/10.1016/B978-0-444-52272-6.00716-9.

Sargan, John D. 1958. “The estimation of economic relationships using instrumental variables.” *Econometrica*, 26(3): 393–415, DOI: http://dx.doi.org/10.2307/1907619.

Sarkar, Sudipa, Soham Sahoo, and Stephan Klasen. 2019. “Employment transitions of women in India: A panel analysis.” *World Development*, 115: 291–309, DOI: http://dx.doi.org/10.1016/j.worlddev.2018.12.003.

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, DOI: http://dx.doi.org/10.1080/00220388.2015.1056784.

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, DOI: http://dx.doi.org/10.1093/aje/kwp100.

Staiger, Douglas O., and James H. Stock. 1997. “Instrumental variables regression with weak instruments.” *Econometrica*, 65(3): 557–586, DOI: http://dx.doi.org/10.2307/2171753.

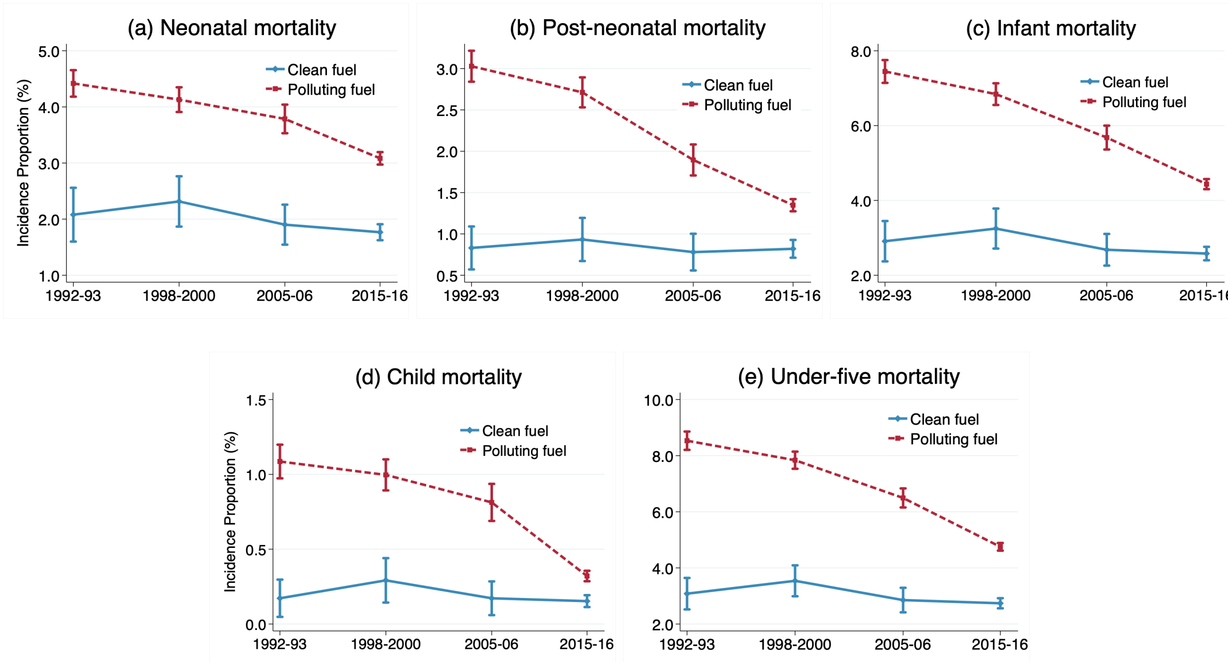
Stock, James H., and Motohiro Yogo. 2005. “Testing for weak instruments in linear IV regression.” In *Andrews DWK Identification and Inference for Econometric Models*. New York, NY: Cambridge University Press, 80–108, DOI: http://dx.doi.org/10.1017/CBO9780511614491.006.

Stoner, Oliver, Jessica Lewis, Itzel Lucio Martínez, Sophie Gumy, Theo Economou, and Heather Adair-Rohani. 2021. “Household cooking fuel estimates at global and country level for 1990 to 2030.” *Nature Communications*, 12(1): 1–8, DOI: http://dx.doi.org/10.1038/s41467-021-26036-x.

Van Kippersluis, Hans, and Cornelius A. Rietveld. 2018. “Beyond plausibly exogenous.” *The Econometrics Journal*, 21(3): 316–331, DOI: http://dx.doi.org/10.1111/ectj.12113.

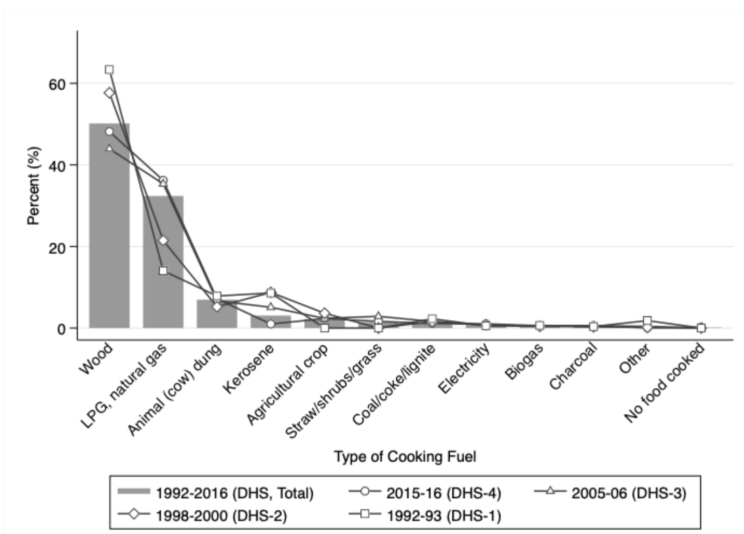
WHO. 2021. “Air quality and health.” https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/sectoral-interventions/household-air-pollution/strategies, Accessed: 2021-11-16.

Figure 1: Mortality trends in different child's age groups by fuel type in India



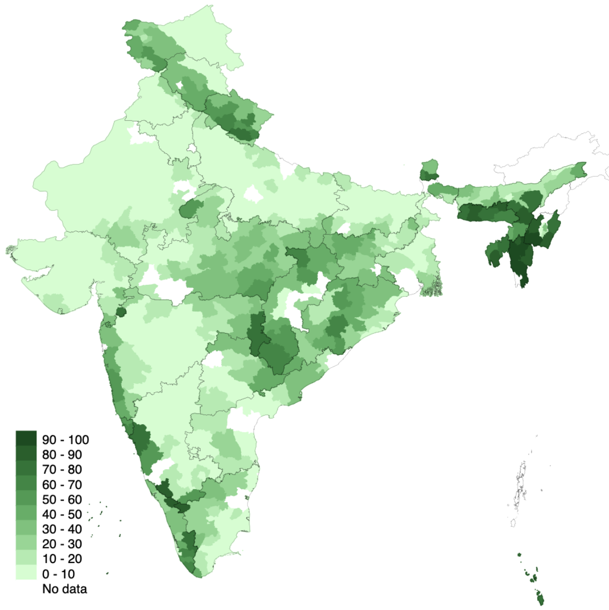
Notes: Based on DHS datasets 1992-1993, 1998-2000, 2005-2006, and 2015-2016. In the medical literature, the measure of incidence proportion (or cumulative incidence) is described as the fraction of children alive at the start of a period who die over that period (Greenland and Rothman, 2008; Centers for Disease Control and Prevention, 2006). The sample weights are applied to adjust the estimates according to the sampling design. Notice that the incidence proportions of neonatal and post-neonatal mortality add up to infant mortality incidence because (i) these two preceding and successive age groups fully make up the first year of life, and (ii) the measure of mortality incidence for all three different age groups have been calculated using a common denominator (or total number of live births). Similarly, the incidence proportions of infant and child mortality add up to under-five mortality incidence.

Figure 2: Share of households in the DHS relying on different types of fuels for cooking



Notes: The figure shows the share of households covered in four rounds of Demographic and Health Survey (DHS) 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 round of the survey, while the bar chart illustrates the share for all four rounds of the survey between 1992 and 2016 (the bars for clean fuels are filled with pattern, whereas the bars for polluting fuels are in solid fill).

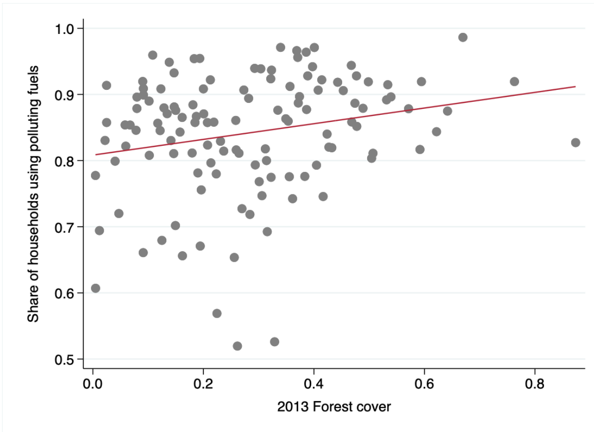
Figure 3: India's district-wise and satellite-based forest cover (2013)



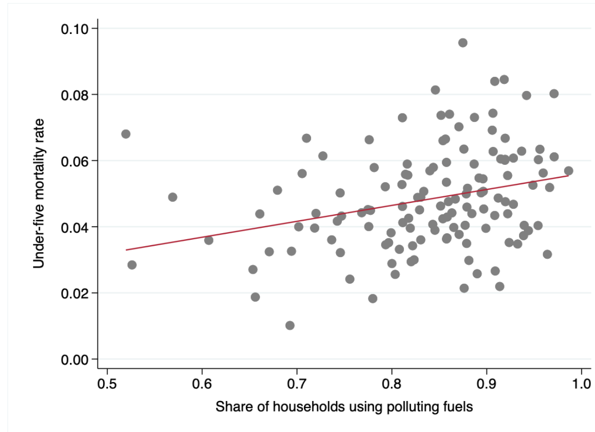
Notes: Based on satellite-based data on forest cover from the Planning Commission of India. The figure depicts the 2013 district-wise forest cover (measured by the percentage of the geographical area covered by forests) in India. The forest cover includes all types of forests (different canopy density 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 of less than 10% are not considered for calculating forest cover.

Figure 4: Relationship between forest cover, cooking fuel choice, and under-five mortality

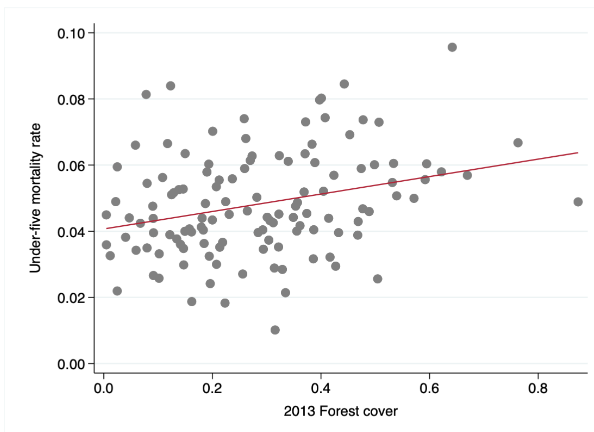
(a) Forest cover and polluting fuel use



(b) Polluting fuel use and under-five mortality



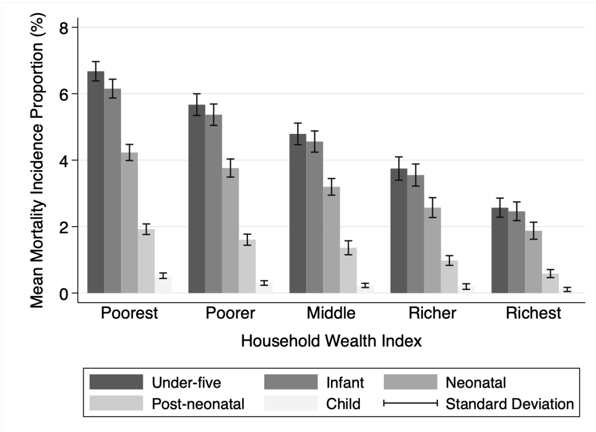
(c) Forest cover and under-five mortality



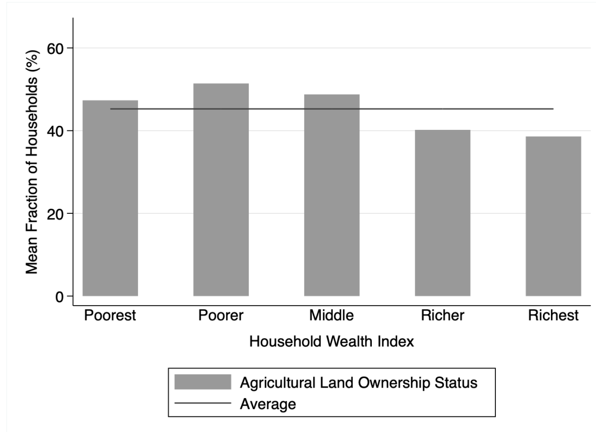
Notes: Based on DHS-4 (2015-2016) and the satellite-based data on forest cover data for 2013. The figure plots the relationship between forest cover, cooking fuel choice, and under-five mortality at the district level for the five states of Jharkhand, Odisha, Assam, Chhattisgarh, and Madhya Pradesh. The child and household-level indicators for cooking fuel choice and mortality are aggregated at the district level using survey weights.

Figure 5: Age-specific infant mortality and ownership status of agricultural land across income distribution in India

(a) Infant mortality



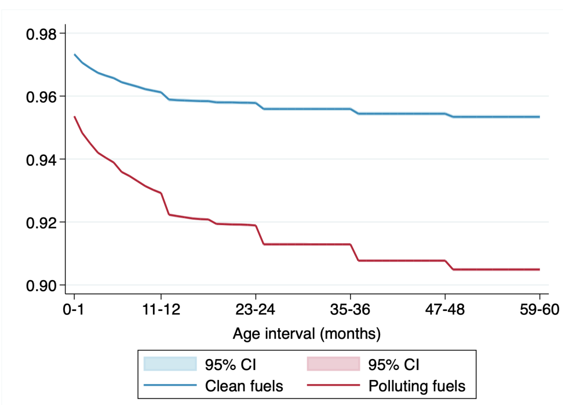
(b) Ownership of agricultural land



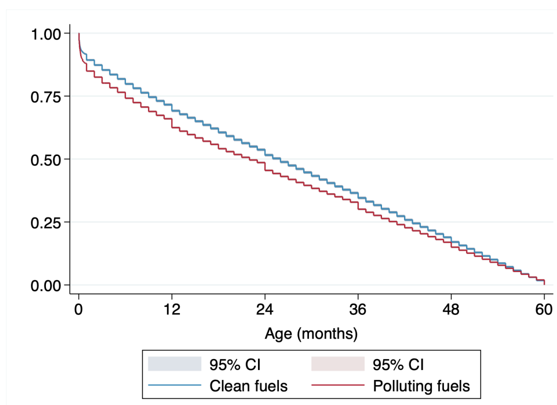
Notes: Based on DHS datasets 1992-1993, 1998-2000, 2005-2006, and 2015-2016. Panel (a) shows that the under-five mortality incidence proportion is higher in lower-income households for household wealth quintiles, suggesting the probability of mortality decreases as a family becomes wealthier. Panel (b) depicts the mean fraction of households that own land for agricultural purposes by dividing agricultural households similarly into quintiles based on household wealth.

Figure 6: Survival analysis

(a) Proportion surviving



(b) Kaplan–Meier survival estimate



Notes: Based on DHS datasets 1992-1993, 1998-2000, 2005-2006, and 2015-2016. Panel (a) depicts the child's proportion of surviving in different age intervals by the choice of household's cooking fuel from the life table using child-level data. Panel (b) presents the Kaplan-Meier estimates on children's survival over age for households using clean and polluting fuels for cooking.

Table 1: Summary of related studies using different study designs

Study	Source of pollution	Health outcomes	Context	Identification	Findings
Panel A. Experimental approach					
Diaz et al. (2007)	Open fire	Women's eye discomfort, headache, and back pain	Rural San Marcos, Guatemala	RCT: Improved stoves (planchas) program	Symptoms of sore eyes, headache, and back pain reduced during the study period of 18 months
Smith-Sivertsen et al. (2009)	Open fire	Women's respiratory symptoms and lung functioning	Rural San Marcos, Guatemala	RCT: Improved stoves (planchas) program	Prevalence of wheezing decreased, but the treatment has no impact on other respiratory symptoms (incl., cough, chronic cough, phlegm, chronic phlegm, wheeze, tightness in chest) and lung functioning
Hanna et al. (2016)	Traditional stove	Lung functioning and various health outcomes among primary cooks (women) and children	Orissa, India	RCT: Improved stove program	No significant impact on health outcomes such as sore eyes, headache, wheeze, and others
Barron and Torero (2017)	Solid fuels	Acute respiratory infections among children under six	El Salvador	RCT: Electrification program	Acute respiratory infections (ARI) was 8-14 percentage points lower among children from encouraged households
Panel B. Quasi-experimental approach					
Pitt et al. (2006)	Hours spent cooking	Adults' and children's respiratory symptoms	Rural Bangladesh and Rural India	IV: Gender-specific hierarchies	10.8 pp. increase per 4 hours for adults and some adverse impact for children
Edwards and Langpap (2012)*	Firewood use, Whether mother cooks	Children's respiratory health	Guatemala	IVs: Gas stove ownership, mother's age	Children living in households that use more wood are more likely to have symptoms of respiratory infection
Silwal and McKay (2015)*	Firewood instead of kerosene, LPG and electricity	Individual's lung capacity damage	Indonesia	IV: Proximity to nearest market	9.4% decrease in individual's lung capacity
Imelda (2018, 2020)	Kerosene and firewood instead of LPG	Infant mortality and child's birth weight	Indonesia	Staggered DID: Fuel-switching program	Switching to LPG reduces infant mortality and birth weight
Liu et al. (2020)	Non-solid vs. Solid fuels	Elderly's ability to handle daily living	Rural China	IV: Share of village citizens who use clean fuels	5.35% and 9.50% increase in activities of daily living and instrumental activities of daily living, respectively, due to non-solid fuel use
Present study	Biomass and other polluting fuels	Under-five mortality by age groups and household size	India	IVs: Forest cover, agricultural land ownership	The estimated effect on under-five, infant, and neonatal mortality rates are 0.040, 0.041, and 0.030. The estimated effect on neonatal mortality rate in households with 5-6 family members is 0.085. Mortality effect is concentrated on girls rather than boys

Notes: Studies with an asterisk (*) find weak justification in the tests to validate the exclusion restriction and exogeneity of their instruments.

Table 2: Testing for positive 2SLS weights

	Polluting fuel use (1)	Owns agricultural land (2)
	Panel A. Continuous measure of forest cover	
Forest cover	0.063***	-0.033
	(0.023)	(0.035)
Owns agricultural land	0.057***	1.000
	(0.003)	—
	Panel B. Binary measure of forest cover	
Forest cover (above mean)	0.023**	-0.022
	(0.009)	(0.013)
Owns agricultural land	0.057***	1.000
	(0.003)	—
p-value: positive weights	1.000	
p-value: negative weights	0.000	

Notes: The table displays regressions of the variable listed in each column on the variable listed on each row. All regressions control for baseline covariates. Standard errors clustered at the district level are in parentheses. In panel A, forest cover is a continuous variable, i.e., district-level percentage of forested area in the total geographical area of the region using satellite-based data. In panel B, forest cover is expressed as a binary variable, specifically, taking a value of 1 if the district's percentage of forest cover is above the mean and 0 otherwise. The first p-value comes from a test of the null hypothesis that the 2SLS weights are all positive, and the second comes from a test of the null hypothesis that at least one weight is negative.

Table 3: Summary statistics

	Mean	SD	Min	Max	Observations
Infant mortality (% total live births)					
Under-five (0-5 years)	0.054	0.225	0	1	360850
Child (1-5 years)	0.005	0.070	0	1	360850
Infant (0-1 year)	0.049	0.215	0	1	360850
Post-neonatal (28 days-1 year)	0.017	0.129	0	1	360850
Neonatal (0-28 days)	0.032	0.176	0	1	360850
Type of cooking fuel (clean)	0.254	0.435	0	1	345932
Place of residence (urban)	0.256	0.436	0	1	360850
Number of household members	6.941	3.378	1	46	360850
Child's gender (female)	0.480	0.500	0	1	360850
Child's breastfeeding status (ever)	0.652	0.476	0	1	360850
Household wealth (wealth index)					
High	0.145	0.352	0	1	360850
Middle	0.384	0.486	0	1	360850
Low	0.472	0.499	0	1	360850
Place where food is cooked					
In same room as they live in	0.368	0.482	0	1	245108
In separate kitchen inside the house	0.437	0.496	0	1	245108
In a separate building	0.099	0.299	0	1	245108
Outdoors	0.097	0.296	0	1	245108
Type of house					
Pucca	0.403	0.490	0	1	349894
Semi-pucca	0.395	0.489	0	1	349894
Kachha	0.202	0.401	0	1	349894
Mother's age (years)					
40-49	0.023	0.149	0	1	360850
<20	0.049	0.216	0	1	360850
20-29	0.698	0.459	0	1	360850
30-39	0.231	0.421	0	1	360850
Mother's education					
Secondary/Higher	0.452	0.498	0	1	360653
Primary	0.145	0.352	0	1	360653
No education	0.403	0.491	0	1	360653
Adult mortality rate (%, district-wise)					
15-49 years old	0.010	0.003	0.000	0.027	301114
50+ years old	0.085	0.015	0.023	0.130	301114

Notes: The table summarizes the household and individual characteristics of respondents from the three rounds of DHS (1992-1993, 1998-2000, and 2015-2016) used in the regressions. The unit of observation is the child, and sampling weights are applied. Neonatal = first 28 days after birth, Post-neonatal = period between the first 28 days after birth and end of the first year of life, Infant = first year of birth, and Child = period from age of one to five. Mean under-five mortality (the sum of mean child and infant mortality) exceeds any individual components that are subsets of a total number of incidents in the first five years of life. Similarly, infant mortality equals the sum of mean post-neonatal and neonatal mortality because these two preceding age groups make up the infant period. 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 4: Summary statistics of infant mortality and fuel choice (by states)

Panel A. Infant mortality (fraction)

States	Under-five (0-5 years)		Child (1-5 years)		Infant (0-1 year)		Post-neonatal (28 days-1 year)		Neonatal (0-28 days)		Obs.
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Uttar Pradesh	0.077	0.267	0.007	0.084	0.070	0.255	0.025	0.155	0.045	0.208	56090
Madhya Pradesh	0.070	0.255	0.008	0.089	0.062	0.241	0.023	0.149	0.039	0.194	31171
Assam	0.062	0.242	0.008	0.091	0.054	0.226	0.020	0.141	0.033	0.180	14015
Bihar	0.061	0.239	0.005	0.073	0.055	0.228	0.018	0.132	0.037	0.190	32580
Chhattisgarh	0.060	0.238	0.006	0.075	0.054	0.227	0.015	0.123	0.039	0.194	9826
Rajasthan	0.058	0.234	0.005	0.070	0.053	0.225	0.021	0.143	0.033	0.178	24964
Odisha	0.058	0.233	0.004	0.064	0.054	0.225	0.021	0.143	0.033	0.178	16192
Delhi	0.055	0.228	0.004	0.065	0.051	0.219	0.022	0.146	0.029	0.167	3700
Jharkhand	0.050	0.218	0.004	0.062	0.046	0.210	0.014	0.119	0.032	0.175	13427
Andhra Pradesh	0.049	0.216	0.003	0.054	0.046	0.210	0.017	0.130	0.029	0.168	5515
All States/UTs	0.054	0.225	0.005	0.070	0.049	0.215	0.017	0.129	0.032	0.176	360850

Panel B. Type of cooking fuel (fraction)

States	Mean		SD	Obs.
	Polluting	Clean		
Bihar	0.906	0.094	0.292	31086
Jharkhand	0.896	0.104	0.306	12712
Meghalaya	0.888	0.112	0.315	6247
Odisha	0.885	0.115	0.319	15487
Assam	0.867	0.133	0.339	13860
Chhattisgarh	0.853	0.147	0.354	9280
West Bengal	0.844	0.156	0.363	9033
Nagaland	0.828	0.172	0.377	5646
Tripura	0.825	0.175	0.380	2500
Madhya Pradesh	0.813	0.187	0.390	29836
All States/Uts	0.746	0.254	0.435	345932

Notes: The table summarizes the infant mortality of five different age groups (outcome variables, top panel) and the type of cooking fuel (key explanatory variable, bottom panel) by state recorded in three rounds of DHS (1992-1993, 1998-2000, and 2015-2016) used in the regressions. The sampling weights are applied. All 35 regions of India (29 states and six union territories–UTs) are considered.

Table 5: Probit: Marginal impact of cooking fuel choice on under-five mortality

	Dependent variable: Child mortality for various age-groups		
	(1)	(2)	(3)
	Panel A. Under-five (0-5 years)		
Polluting fuel for cooking	0.023***	0.008***	0.008***
	(0.001)	(0.001)	(0.001)
Observations	345,932	221,913	221,913
Probit log-likelihood	-69,858	-35,657	-35,647
	Panel B. Child (1-5 years)		
Polluting fuel for cooking	0.0040***	0.0008**	0.0008**
	(0.0004)	(0.0004)	(0.0004)
Observations	343,593	219,861	219,861
Probit log-likelihood	-10,264	-3,923	-3,923
	Panel C. Infant (0-1 year)		
Polluting fuel for cooking	0.020***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)
Observations	345,932	221,913	221,913
Probit log-likelihood	-65,227	-33,909	-33,901
	Panel D. Post-neonatal (28 days-1 year)		
Polluting fuel for cooking	0.009***	0.001	0.001
	(0.001)	(0.001)	(0.001)
Observations	345,932	221,518	221,518
Probit log-likelihood	-293,29	-13,840	-13,839
	Panel E. Neonatal (0-28 days)		
Polluting fuel for cooking	0.011***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)
Observations	345,932	221,913	221,913
Probit log-likelihood	-46,777	-25,420	-25,411
Year FE	Yes	Yes	No
State FE	Yes	Yes	No
Demographic controls	No	Yes	Yes
State × Year FE	No	No	Yes

Notes: Each column reports AMEs for probit regression where the key explanatory variable is polluting fuel for cooking. The dependent variable is child mortality for different age groups: under-five, child, infant, post-neonatal, and neonatal in panels A through E, respectively. The year fixed effects in Columns (2) and (3) include dummies for two years of interviews (2015 and 2016). The state fixed effects include dummies for 36 states. 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; infant characteristics: gender and breastfeeding status; and district characteristics: age-specific adult mortality rates. 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: Effect of cooking fuel choice on infant mortality from IV regressions
(IV = Agricultural land ownership)

	1^{st} stage	2^{nd} stage				
	(1)	(2)	(3)	(4)	(5)	(6)
	Polluting fuel use	Under-five (0-5 years)	Child (1-5 years)	Infant (0-1 year)	Post-neonatal (28 days-1 year)	Neonatal (0-28 days)
Polluting fuel for cooking		0.047***	0.001	0.046***	0.010	0.036***
		(0.018)	(0.004)	(0.017)	(0.010)	(0.014)
Owns agricultural land	0.057***					
	(0.002)					
Observations	221913	221913	221913	221913	221913	221913
R^2	0.53	0.02	0.00	0.02	0.01	0.01
Montiel Olea-Pflueger weak IV test						
Effective F-statistic ($\alpha = 5\%$)	772.33					
Critical value 2SLS ($\tau = 10\%$)	23.11					

Notes: 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; infant characteristics: gender and breastfeeding status; and district characteristics: age-specific adult mortality rates. The state-by-year fixed effects 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 five 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 7: Effect of cooking fuel choice on infant mortality from IVs regressions
(IVs = Forest cover and agricultural land ownership)

	1^{st} stage	2^{nd} stage				
	(1)	(2)	(3)	(4)	(5)	(6)
	Polluting/Biomass fuel use	Under-five (0-5 years)	Child (1-5 years)	Infant (0-1 year)	Post-neonatal (28 days-1 year)	Neonatal (0-28 days)
		Panel A. Indoor air pollution = Polluting fuel				
Polluting fuel for cooking		0.040**[**]	-0.001	0.041**[**]	0.011	0.030**[**]
		(0.020)	(0.005)	(0.019)	(0.011)	(0.015)
		[0.018]	[0.005]	[0.017]	[0.010]	[0.014]
Forest cover	0.065***					
	(0.022)					
Owns agricultural land	0.054***					
	(0.003)					
Observations	194254	194254	194254	194254	194254	194254
R^2	0.53	0.02	0.00	0.02	0.01	0.01
Montiel Olea-Pflueger weak IV test						
Effective F-statistic ($\alpha = 5\%$)	54.12					
Critical value 2SLS ($\tau = 10\%$)	17.41					
Hansen J statistic		1.70	0.33	2.18	0.40	2.29
Degree of overidentification		1.00	1.00	1.00	1.00	1.00
p-value of Hansen J statistic		0.19	0.57	0.14	0.53	0.13
		Panel B. Indoor air pollution = Biomass fuel				
Biomass fuels for cooking		0.037*[**]	-0.001	0.037**[**]	0.011	0.026*[**]
		(0.019)	(0.005)	(0.018)	(0.010)	(0.014)
		[0.015]	[0.004]	[0.015]	[0.008]	[0.012]
Forest cover	0.067***					
	(0.022)					
Owns agricultural land	0.058***					
	(0.003)					
Observations	189384	189384	189384	189384	189384	189384
R^2	0.55	0.02	0.00	0.02	0.01	0.01
Montiel Olea-Pflueger weak IV test						
Effective F-statistic ($\alpha = 5\%$)	57.82					
Critical value 2SLS ($\tau = 10\%$)	17.74					
Hansen J statistic		1.74	0.58	2.40	0.44	2.54
Degree of overidentification		1.00	1.00	1.00	1.00	1.00
p-value of Hansen J statistic		0.19	0.45	0.12	0.51	0.11

Notes: The first column reports the result from the first-stage regression of our IV (2SLS) regression using DHS-4 data. The dependent variable is a binary variable of whether fuel choice: polluting fuel (top panel) and biomass fuel (bottom panel). The effective F-statistic on IVs—2013 district-wise forest cover calculated as a percentage of forested area in the 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 (top panel) and biomass (bottom panel) fuels for cooking. Columns (2)–(6) 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 (top panel) or biomass fuel (bottom panel) from the first-stage estimation. The number of observations decreases in the bottom panel as we drop non-biomass polluting fuels from the sample to contrast the effect of using polluting biomass fuels to that of using polluting fuels. The Hansen's J-statistic suggests that the excluded IVs are exogenous. 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; infant characteristics: gender and breastfeeding status; and district characteristics: age-specific adult mortality rates. The state-by-year fixed effects are also controlled in every specification. Unit of observation: child. Heteroskedasticity-robust 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 brackets. The statistical significance of the key regressors is 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 8: Heterogeneous mortality effects of IAP by child's age and household size

	1^{st} stage	2^{nd} stage				
	(1)	(2)	(3)	(4)	(5)	(6)
	Polluting fuel use	Under-five	Child	Infant	Post-neonatal	Neonatal
		Panel A. Number of household members = [1-4]				
Polluting fuel for cooking		0.221***	0.011	0.210***	0.082**	0.127**
		(0.066)	(0.017)	(0.064)	(0.035)	(0.051)
Forest cover	0.082***					
	(0.025)					
Owns agricultural land	0.035***					
	(0.004)					
Observations	48777	48777	48777	48777	48777	48777
R^2	0.58	-0.00	0.00	-0.00	-0.01	0.01
Montiel Olea-Pflueger weak IV test						
Effective F-statistic ($\alpha = 5\%$)	22.87					
Critical value 2SLS ($\tau = 10\%$)	14.65					
Hansen J statistic		5.08	0.58	6.49	5.46	2.98
Degree of overidentification		1.00	1.00	1.00	1.00	1.00
p-value of Hansen J statistic		0.02	0.44	0.01	0.02	0.08
		Panel B. Number of household members = [5-8]				
Polluting fuel for cooking		0.059**	0.002	0.056**	0.008	0.048***
		(0.023)	(0.006)	(0.022)	(0.012)	(0.018)
Forest cover	0.070***					
	(0.022)					
Owns agricultural land	0.054***					
	(0.003)					
Observations	109747	109747	109747	109747	109747	109747
R^2	0.53	0.01	0.00	0.01	0.00	0.00
Montiel Olea-Pflueger weak IV test						
Effective F-statistic ($\alpha = 5\%$)	48.60					
Critical value 2SLS ($\tau = 10\%$)	16.57					
Hansen J statistic		1.96	0.03	2.12	0.31	2.20
Degree of overidentification		1.00	1.00	1.00	1.00	1.00
p-value of Hansen J statistic		0.16	0.86	0.15	0.58	0.14
		Panel C. Number of household members = [9-12]				
Polluting fuel for cooking		-0.023	-0.001	-0.022	-0.020	-0.002
		(0.036)	(0.008)	(0.035)	(0.018)	(0.030)
Forest cover	0.021					
	(0.032)					
Owns agricultural land	0.069***					
	(0.007)					
Observations	28278	28278	28278	28278	28278	28278
R^2	0.52	0.01	0.00	0.01	0.00	0.01
Montiel Olea-Pflueger weak IV test						
Effective F-statistic ($\alpha = 5\%$)	37.02					
Critical value 2SLS ($\tau = 10\%$)	10.41					
Hansen J statistic		0.06	0.40	0.01	0.92	0.62
Degree of overidentification		1.00	1.00	1.00	1.00	1.00
p-value of Hansen J statistic		0.80	0.53	0.92	0.34	0.43

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 covered in DHS-4 data. Panel A–C considers each of the three subsamples in an order of 1-4 to 9-12 members. 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)–(6) report results from the estimation of equation (1) using IV regression with different dependent variables and similar specifications. The outcome variable in an IV regression is a binary variable of infant mortality for each of the five different age groups. All specifications contain an unreported vector of demographic controls, state-by-year fixed effects, and a constant term. Heteroskedasticity-robust standard errors clustered by districts are in parentheses. Significance: $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.

Table 9: Heterogeneous mortality effects of IAP by child's age and gender

	1^{st} stage	2^{nd} stage				
	(1) Polluting fuel use	(2) Under-five	(3) Child	(4) Infant	(5) Post-neonatal	(6) Neonatal
		Panel A. Boys				
Polluting fuel for cooking		0.009	-0.006	0.015	-0.005	0.021
		(0.027)	(0.006)	(0.026)	(0.014)	(0.022)
Forest cover	0.058**					
	(0.023)					
Owns agricultural land	0.058***					
	(0.003)					
Observations	101309	101309	101309	101309	101309	101309
R^2	0.54	0.02	0.00	0.02	0.01	0.01
Montiel Olea-Pflueger weak IV test						
Effective F-statistic ($\alpha = 5\%$)	55.64					
Critical value 2SLS ($\tau = 10\%$)	16.66					
Hansen J statistic		1.17	1.18	1.84	0.13	1.92
Degree of overidentification		1.00	1.00	1.00	1.00	1.00
p-value of Hansen J statistic		0.28	0.28	0.17	0.72	0.17
		Panel B. Girls				
Polluting fuel for cooking		0.078***	0.005	0.073***	0.032**	0.041**
		(0.026)	(0.008)	(0.025)	(0.016)	(0.020)
Forest cover	0.072***					
	(0.022)					
Owns agricultural land	0.050***					
	(0.003)					
Observations	92945	92945	92945	92945	92945	92945
R^2	0.53	0.01	0.00	0.01	0.00	0.01
Montiel Olea-Pflueger weak IV test						
Effective F-statistic ($\alpha = 5\%$)	45.27					
Critical value 2SLS ($\tau = 10\%$)	16.29					
Hansen J statistic		1.11	0.17	1.02	0.55	0.48
Degree of overidentification		1.00	1.00	1.00	1.00	1.00
p-value of Hansen J statistic		0.29	0.68	0.31	0.46	0.49

Notes: The table presents heterogeneous treatment effects of IAP on infant mortality by both child's age and gender using the DHS-4 data. Panel A and B are respectively based on a subsample of boys and girls. All specifications contain an unreported vector of demographic controls, state-by-year fixed effects, and a constant term. Heteroskedasticity-robust standard errors clustered by districts are in parentheses. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table 10: Heterogeneous mortality effects of IAP by child's age and status of male siblings

	1^{st} stage	2^{nd} stage				
	(1) Polluting fuel use	(2) Under-five	(3) Child	(4) Infant	(5) Post-neonatal	(6) Neonatal
	Panel A. Lives with male siblings					
Polluting fuel for cooking		0.094*	0.000	0.093*	0.010	0.083*
		(0.053)	(0.013)	(0.051)	(0.030)	(0.044)
Forest cover	0.054**					
	(0.025)					
Owns agricultural land	0.046***					
	(0.004)					
Observations	50366	50366	50366	50366	50366	50366
R^2	0.49	0.02	0.00	0.02	0.01	0.01
Montiel Olea-Pflueger weak IV test						
Effective F-statistic ($\alpha = 5\%$)	34.49					
Critical value 2SLS ($\tau = 10\%$)	13.25					
Hansen J statistic		2.38	0.64	1.97	0.06	2.21
Degree of overidentification		1.00	1.00	1.00	1.00	1.00
p-value of Hansen J statistic		0.12	0.42	0.16	0.81	0.14
	Panel B. Lives without any male siblings					
Polluting fuel for cooking		0.047**	-0.000	0.047***	0.018*	0.029*
		(0.019)	(0.005)	(0.018)	(0.010)	(0.015)
Forest cover	0.068***					
	(0.023)					
Owns agricultural land	0.057***					
	(0.003)					
Observations	143888	143888	143888	143888	143888	143888
R^2	0.54	0.02	0.00	0.02	0.01	0.02
Montiel Olea-Pflueger weak IV test						
Effective F-statistic ($\alpha = 5\%$)	54.56					
Critical value 2SLS ($\tau = 10\%$)	17.61					
Hansen J statistic		1.83	1.25	2.79	1.00	2.02
Degree of overidentification		1.00	1.00	1.00	1.00	1.00
p-value of Hansen J statistic		0.18	0.26	0.09	0.32	0.16

Notes: The table presents heterogeneous treatment effects of IAP on infant mortality by both child's age and the status of male siblings using the DHS-4 data. Panel A and B are respectively based on a subsample of children with and without male siblings. All specifications contain an unreported vector of demographic controls, state-by-year fixed effects, and a constant term. Heteroskedasticity-robust standard errors clustered by districts are in parentheses. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table 11: Regression results from the zero-first-stage test

	(1) IV = Agricultural land ownership	(2) IVs = Agricultural land ownership and forest cover
	Effect of instrument(s) on polluting fuel use for cooking	
	Panel (a). Reduced form (full sample)	
Owns agricultural land	0.1070***	0.1050***
	(0.0042)	(0.0044)
Forest cover		0.0899***
		(0.0232)
Montiel Olea-Pflueger weak IV test		
Effective F-statistic ($\alpha = 5\%$)	636.41	177.53
Critical value 2SLS ($\tau = 10\%$)	23.11	12.01
Observations	220,572	194,254
	Panel (b). Direct effect (zero-first-stage group)	
Owns agricultural land	0.0047	-0.0053
	(0.0145)	(0.0146)
Forest cover		0.0619
		(0.0460)
Montiel Olea-Pflueger weak IV test		
Effective F-statistic ($\alpha = 5\%$)	0.10	1.21
Critical value 2SLS ($\tau = 10\%$)	23.11	9.92
Observations	4,906	4,287
	Panel (c). Direct effect (remaining sample)	
Owns agricultural land	0.1048***	0.1028***
	(0.0042)	(0.0044)
Forest cover		0.0895***
		(0.0229)
Montiel Olea-Pflueger weak IV test		
Effective F-statistic ($\alpha = 5\%$)	613.81	175.91
Critical value 2SLS ($\tau = 10\%$)	23.11	11.66
Observations	215,666	189,967
	Effect of instrument(s) on under-five mortality	
	Panel (d). Reduced form (full sample)	
Owns agricultural land	0.0020*	0.0021*
	(0.0010)	(0.0011)
Forest cover		-0.0021
		(0.0032)
Observations	220,572	194,254
	Panel (e). Direct effect (zero-first-stage group)	
Owns agricultural land	-0.0043	-0.0098
	(0.0106)	(0.0117)
Forest cover		0.0275
		(0.0280)
Observations	4,906	4,287
	Panel (f). Direct effect (remaining sample)	
Owns agricultural land	0.0027**	0.0028**
	(0.0010)	(0.0011)
Forest cover		-0.0027
		(0.0032)
Observations	215,666	189,967

Notes: All specifications include an unreported constant term, state-by-year fixed effects, and baseline demographic controls except for urban/rural dummy. Heteroskedasticity-robust standard errors clustered by districts are in parentheses. Significance: $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.