American Economic Journal: Economic Policy 2016, 8(1): 80–114
http://dx.doi.org/10.1257/pol.20140008

Up in Smoke: The Influence of Household Behavior on the Long-Run Impact of Improved Cooking Stoves†

By Rema Hanna, Esther Duflo, and Michael Greenstone*

Laboratory studies suggest that improved cooking stoves can reduce indoor air pollution, improve health, and decrease greenhouse gas emissions in developing countries. We provide evidence, from a large-scale randomized trial in India, on the benefits of a common, laboratory-validated stove with a four-year follow-up. While smoke inhalation initially falls, this effect disappears by year two. We find no changes across health outcomes or greenhouse gas emissions. Households used the stoves irregularly and inappropriately, failed to maintain them, and usage declined over time. This study underscores the need to test environmental technologies in real-world settings where behavior may undermine potential impacts. (*JEL* D12, O12, O13, Q53, Q54, Q55)

A third of the world's population, and up to 95 percent in poor countries, rely on solid fuels, including biomass (e.g., wood, dung, agricultural residues) and coal, to meet their energy needs (Gordon et al. 2014). The World Health Organization lists "indoor air pollution (IAP) from primitive household cooking fires as the leading environmental cause of death in the world," estimating that it was responsible for 4.3 million deaths annually, or about as many deaths as malaria and tuberculosis combined. Moreover, cooking with biomass fuels is a key source of climate change through its release of carbon dioxide (CO_2) and black carbon (Kandlikar, Reynolds, and Grieshop 2009). In response, improved cooking stoves are increasingly seen as

* Hanna: Harvard University, 79 JFK Street, Cambridge, MA 02138 and National Bureau of Economic Reseearch (NBER) and Abdul Latif Jameel Poverty Action Lab (J-PAL) (e-mail: rema_hanna@hks.harvard.edu); Duflo: Massachusetts Institute of Technology (MIT), 77 Massachusetts Avenue, Building E17, Room 201B, Cambridge, MA 02139 and NBER and J-PAL (e-mail: eduflo@mit.edu); Greenstone: University of Chicago, 1126 E. 59th Street, Chicago, IL 60637 and NBER and J-PAL (e-mail: mgreenst@uchicago.edu). This project is a collaboration involving many people and organizations. Foremost, we are deeply indebted to Gram Vikas and especially to Joe Madiath, who made this research possible. We are grateful for insightful comments from Jessica Cohen, Pascaline Dupas, Edward Glaeser, Seema Jayachandran, Margaret McConnell, Grant Miller, Mushfiq Mobarak, Rohini Pande, and Rebecca Thornton and seminar participants at Harvard University, University of Michigan Michigan, University of California, San Diego, and the NBER Environmental Economics Meetings. We thank the four anonymous referees who helped make this a better paper. We thank John McCracken for advice on emissions monitoring and Dr. Vandana Sharma for training the team on health monitoring. We thank Yusuke Taishi, Raymond Guiteras, Ritwik Sakar, Annie Duflo, Reema Patnaik, Anup Kumar Roy, Shobhini Mukerji, Mihika Chatterjee, Trevor Bakker, and KB Prathap for their excellent work coordinating the fieldwork. Sarah Bishop, Gabriel Tourek, Mahvish Shaukat, Samuel Stopler, and Francine Loza provided superb research assistance. For financial support, we thank the MIT Energy Initiative, the Centre for Microfinance at the Institute of Financial Management and Research, the Institut Veolia Environnement, and the Children's Investment Fund Foundation. A portion of this work was conducted while Dr. Hanna was a fellow at the Science Sustainability Program at Harvard University.

† Go to http://dx.doi.org/10.1257/pol.20140008 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

a tool to improve respiratory health and combat climate change.[1] For example, in September 2010, Hillary Clinton announced the formation of the Global Alliance for Clean Cookstoves (GACC), which calls for 100 million homes to adopt clean and efficient stoves and fuels by 2020. This big push for improved cooking stoves has occurred despite surprisingly little rigorous evidence on their efficacy on health and fuel use in real-life settings.[2]

As is the case with other preventive health interventions, the key gap in knowledge lies in the translation between technological efficacy, which is established in laboratory-like conditions, and effectiveness given "typical" use.[3] For improved stoves, effectiveness depends on modifications to household behavior to ensure their proper use and maintenance. Recent evidence (Mobarak et al. 2012) suggests a very low willingness to pay for improved stoves, and little interest in their health impacts among women and, especially, men. If households are not willing to pay the upfront cost of acquiring an improved stove and do not seem concerned about the problem they seek to address, it is conceivable that they would not be willing to pay the small, but recurring, everyday costs that are associated with switching from their traditional technology and maintaining the improved stove.

This problem is one of independent policy interest, given the enormous resources that have been committed to stove distribution, as well as the health consequences of indoor air pollution and the climate change implications. It is also a specific instance of a very generic issue: there is a strong belief in policy circles that technological solutions are a key part of the solution to many development problems (for example, Bill Gates' effort to "reinvent the toilet").[4] However, in many cases (e.g., traditional bed nets that need to be both used and regularly re-treated with insecticide, cell-phone based savings accounts where money needs to be put in more than once, antibiotic courses that need to be finished, water filters that need to be used regularly, or improved water sources that households need to switch to and maintain), a technological innovation requires a complementary investment from the households, which needs to be sustained over time. When the widespread belief in the value of the technology (i.e., protection against malaria, savings, health, or clean water) is low, can we expect the households to sustain this behavior change over time?[5]

[1] These benefits are cited regularly in leading publications, including *The New York Times*, and they have a range of proponents from Bill Clinton to Julia Roberts.

[2] According to the GACC website, the stoves lead to dramatic reductions in child pneumonia; save the equivalent of 1–2 tons of CO_2 per year; and produce fuel savings that families can use to pay for the stove. However, as we discuss, none of the evidence, to date, fully supports these claims.

[3] For example, key randomized evaluations of bed nets (Alonso et al. 1991; Phillips-Howard et al. 2003; Binka, Indome, and Smith 1998; Nevill et al. 1996) sent project staff to re-treat the nets every six months, where presumably households are also reminded to use them. In Phillips-Howard et al. (2003), households signed forms that the nets remained the property of the project until after the study was concluded, which could have induced households to keep the nets in better condition than if they had been told that the nets were their own. This issue extends beyond the developing world: Duggan (2005) argues that in the United States antipsychotic drugs are less effective in practice than in FDA trials due to several factors, including short-run follow-ups and prescribing the drugs for people that differ from the individuals enrolled in the clinical trials.

[4] "Bill Gates Can't Build a Toilet," *The New York Times*, November 18, 2013.

[5] Some of these devices are designed with the goal to make it cheaper for households to keep up with the new behavior, and they may subsequently learn from use the value of what they bring. For example, Dupas (2014) studies long-lasting insecticide treated bednets, which do not need to be re-treated and are much more comfortable to

This paper provides experimental evidence on the links between the distribution of almost free improved cooking stoves and individual behavior and well-being, and greenhouse gas emissions in Orissa, India. Gram Vikas (GV), an award winning nongovernmental organization (NGO), obtained funding to subsidize stove construction for 15,000 households over five years, independent of this research. GV chose stoves designed with an enclosed cooking chamber (to keep the flame separate from the food) and a chimney to direct smoke away from the user. The stoves had been proven to reduce IAP and energy consumption in laboratory settings and could be constructed with locally sourced materials, facilitating distribution at a large scale. The laboratory tests were replicated in GV's lab. At a total cost of about US$12.50, these stoves fall within the "lower end" of improved stove technologies and, since the start of the study, there have been improvements in their design. However, these stoves represent the vast majority of improved stoves that have been distributed to date: the World Bank (2011) reports that stove programs have typically distributed improved stoves in this category and over 166 million of them are in use today. GV designed a rollout strategy with considerable thought and attention as to how to advertise the benefits of the stoves, train households to use the stoves properly, and provide support for maintenance and repairs. Thus, considering both the successes of their other programs (Duflo et al. 2014), as well as the way they structured the stove rollout, this is a "better than typical" rollout for an NGO given the funding level and large scope of the program.

A public lottery determined the order in which stoves were constructed within each village for 2,600 households. The first third of households within each village received the stoves at the start of the project, the second third received the stoves about two years after the first wave, and the remaining households received them at the end. Households were followed for four years after the initial stove offers, which allowed for an examination of the long-run use and impacts of the stoves. This long-run follow-up is relatively rare in the evaluations of health interventions or other new technologies where households learn about the benefits and maintenance needs of the product over time.

There are four primary results. First, initial household take-up and usage of the new stoves was far from universal and then declined markedly over time as households failed to make the maintenance investments (e.g., cleaning the chimney) necessary to keep them fully operational. Several measures document this, but perhaps the most salient is that treatment households that received the GV improved stoves still continued to use their traditional stoves in conjunction with the new ones—even early on, when the majority of the stoves were functional. In the early years, treatment households only cooked 3.5 more meals per week (or 25 percent of total meals) with an improved stove in good condition than the control households.[6] This difference was halved to about 1.8 meals per week in year three, as the stoves deteriorated.

sleep under. Subsidizing the fixed cost of a better net helps households use bednets more regularly. Improved cookstoves often claim to be more energy efficient, which would be valued by the households (Mobarak et al. 2012).

[6] Beltramo and Levine (2010) observed the same phenomenon with solar ovens (another type of stove used to reduce smoke exposure and energy consumption) in Senegal; even households that chose to use the solar oven generally cooked only a few of their meals on them, continuing to cook the remaining meals on their standard stoves.

Second and correspondingly, the stoves failed to achieve their primary goal of reducing exposure to hazardous air pollutants. While there was a significant effect on smoke inhalation during the first year for the primary cooks in the household (though not for children), the treatment effect became statistically indistinguishable from zero in subsequent years as usage rates and maintenance declined. Further, even in the first year, the resulting effect (a 7.5 percent decrease in the carbon monoxide concentration of exhaled breath) was smaller than the reduction observed in laboratory-style settings with properly maintained stoves and near-perfect usage rates.

Third, we cannot reject the null hypothesis that the stoves failed to affect health across a wide set of health outcomes. For example, there is no difference in lung functioning (as measured by spirometry tests) between women who regularly cook in the treatment and control groups. Furthermore, we fail to find a positive impact on a wide variety of measured and self-reported health outcomes, including infant birth weight, infant mortality rates, probability of a cough, blood pressure, or even the probability of any illness in the last 30 days. This does not appear to be due to a lack of power.

Fourth, the treatment group appears to have experienced modest *declines* in their living standards and there is no evidence of a reduction in greenhouse gas emissions. Specifically, treatment households spent substantially more time repairing their stoves. Furthermore, the treatment did not affect fuel costs or time spent cooking, which is consistent with the energy consumption results of "health-improving" stoves studied by Miller and Mobarak (2014) and Burwen and Levine (2011). There is also no evidence of potential climate benefits from reductions in deforestation since there was no change in the total amount of wood used for cooking. It is noteworthy that these findings contrast with self-reported satisfaction of improved stoves and laboratory test results that show reduced time and energy used to boil the same quantity of water with an improved stove.

Besides demonstrating the importance of accounting for human behavior in assessing the effectiveness of new technologies, this study builds upon and contributes to the literature on indoor air pollution. This study remains the only experimental evidence of improved stoves' long-run health outcomes under realistic usage conditions. Most evidence on IAP comes from observational studies that compare fuel use and health status of users and nonusers (e.g., Bruce, Perez-Padilla, and Albalak 2000). However, households that cook with improved stoves are typically different in other respects, such as income levels and health preferences, as well (Bruce et al. 1998, Mueller et al. 2011). As a result, it is unclear whether these studies' estimated positive effects of reducing IAP reflect the impact of improved stoves or unobservable characteristics.

It is therefore important to consider experimental evidence. For example, Bensch and Peters (2012) used randomized evaluation techniques and found that an improved stove program caused reductions in self-reported respiratory and eye disease for women in a sample of 227 households in Senegal, implying that there may be returns to use of the improved stoves. Experimental evidence has also emerged from the RESPIRE study, an evaluation of a concrete stove in Guatemala (Smith-Sivertsen et al. 2009).

Our paper complements the literature in at least two important ways. First, we followed households for four years after the receipt of the stoves, a length of time much greater than in any previous study. For example, the RESPIRE study follows households for 12 months for the full sample and 18 months for a small subset, Bensch and Peters (2012) follows households for one year, and Beltramo and Levine (2010) follows households for six months. Our extended evaluation may be important for several reasons. First, the treatment effects on health may change considerably over time, as households learn about the value of the stoves and subsequently change their usage rates and maintenance investments, as well as experience a general depreciation of the technology. Second, the effect on health may be cumulative over several years. Third, we found meaningful effects on carbon monoxide (CO) for primary cooks in the first year; had we ended the study after learning this, we would have projected the effect for several years in benefit-cost calculations. In reality, this effect was very short-lived.[7]

Perhaps most importantly, we study an actual program run by a local nonprofit with little assistance by the research team. The stoves are locally made and relatively cheap (roughly \$12.50), implying that they would be practical for large-scale distribution and presumably affordable for the target population if sold (annual per capita consumption of households in our sample is \$145). By comparison, the stoves in the RESPIRE study cost between \$100 and \$150, making them prohibitively expensive for most households where indoor air pollution is a problem. Furthermore, although the RESPIRE study was conducted in the field, trained fieldworkers inspected the stoves weekly for proper use and maintenance and then arranged for repairs as needed (Smith et al. 2010). In this respect, the RESPIRE study shuts down households' ability to reveal their valuation through usage rates and decisions about shifting resources from other goods to stove maintenance. Thus, the results from the RESPIRE study likely provide upper bound effects, while our estimates more closely resemble real-world impacts, where households may not use the technology appropriately or may choose not to use the technology at all. The mixed results on health from the RESPIRE study (discussed in depth below) and the lack of health impacts found in our study, which derive from limited and improper use, suggest that, in the context of evolving stove technologies, the new generation of stoves (e.g., envirofit and rocket stoves) need to be evaluated in field settings to understand whether their real-world benefits match their laboratory benefits before valuable resources are devoted to their deployment.

More generally, this paper contributes to the literature on the adoption of health and environmental technologies. Many times, new technologies are evaluated in laboratory experiments or through field experiments in controlled settings where researchers ensure high compliance in terms of use and maintenance. These studies are vital because they provide an upper bound effect on the possible treatment effects of the technologies. However, as Chassang, Padró i Miquel, and Snowberg (2012) discuss, perfectly controlling individual choices and actions produces an impact estimate that, although internally valid, lacks external validity when a

[7] Note that our study's sample included over 4,000 women and 3,000 children, compared to the 500 women and children in the RESPIRE study, which provides greater precision in detecting any health and fuel effects.

complementary household action is needed. To account for this, studies like this one, where households are free to adjust their behavior (or not) over the long run, are necessary. There are fewer of these studies: they take a longer time to carry out and lack the crispness of more tightly-controlled studies in terms of identifying unique causal channels (and they are more likely to produce null effects). However, they are vital to the production of knowledge on which policy can actually be based.

The remainder of the paper proceeds as follows. Section I discusses the experimental design. The data is described in Section II, while the empirical framework is laid out in Section III. The findings are presented in Section IV. Section V provides a discussion of the state of knowledge on improved stoves to date, as well as lessons that can be learned for future research and evaluation. Section VI concludes the paper.

I. Experimental Design

A. *Setting*

This project took place in India, where about 70 percent of the population burn solid fuels—firewood, crop residue, or cow dung—in traditional stoves (see Appendix Figure 1, panel A) to meet their cooking needs (Census of India 2001). The reliance on traditional fuels is even higher (90 percent) in poorer, rural regions. Indoor air pollution (IAP) levels from traditional stoves are high. For example, Smith (2000) reports that the "available data show a distribution of indoor PM_{10} 24-h concentrations measured in Indian solid-fuel-using households ranging to well over 2,000 $\mu g/m^3$." To put these figures into context, the Central Pollution Control Board of India states that ambient levels of PM_{10} should not exceed 100 $\mu g/m^3$.

In response to the health threats posed by the use of solid fuel in traditional stoves, as well as concerns about deforestation, both governments and nongovernmental organizations (NGOs) have been implementing clean stove programs for several decades. For example, during the 1980s and 1990s, the Indian government alone subsidized and distributed 32 million improved stoves. However, many of these stoves had life spans of less than two years, and as Smith (2000) has pointed out, only a small fraction of the stoves built before 1990 still existed at the time of his article. In fact, this campaign is widely acknowledged to have been a failure, with the stoves laying unused or rapidly falling into disrepair (Block 2013). The renewed interest in IAP worldwide has prompted a new wave of interest in India as well, with NGOs, local governments, and private foundations investing in the design and distribution of improved stoves, and the launch in 2011 of a new large-scale government program with an improved design.

This paper evaluates an improved stove program run by Gram Vikas, an NGO that operates in the state of Orissa. Orissa is one of the poorest states in India, with 40 percent of the population living below the poverty line. Poverty is significantly worse in the western and southern districts of the state where this project took place. Gram Vikas is considered to be a top nonprofit, having won numerous awards, including being listed in the *Global Journal's* "Top 100 Best NGOs in the World" in 2012. Their sanitation program has been rigorously evaluated and shown to have led to considerable health benefits (see Duflo et al. 2014). Independent of the researchers, Gram Vikas obtained funding from the Inter-Community Church Organization (ICCO) to subsidize the construction of the stoves to roughly 15,000 households over five years.

The stove considered in this study represents a relatively inexpensive improved stove technology. It was developed and tested by the Appropriate Rural Technology Institute (ARTI), an NGO specializing in energy innovation for rural areas. Like the traditional stoves, it is largely made out of mud (see online Appendix Figure 1, panel B). However, the constructed base encloses the cooking flame and it includes a chimney to direct smoke away from the user. Moreover, it allows for two pots, instead of the one pot in traditional stoves, to potentially reduce cooking time.

The chosen stove was considered appealing because it is constructed with local materials and is low-cost at roughly US$12.50. Gram Vikas subsidized the stove cost by contributing stove materials (chimney), design, and a skilled mason to supervise the construction. Households were responsible for providing mud for the stove base, labor and a payment of about US$0.75, which was used both to pay the mason who assisted in building and maintaining the stoves and to contribute to a fund for stoves for any new houses built in the village. As the stove is made from locally available materials, it can be easily constructed in these remote, rural areas of India.

In laboratory settings, the ARTI stove burns more efficiently than a traditional stove, leading to lower biofuel requirements and less indoor smoke. However, obtaining this outcome requires that the stoves are maintained appropriately, which involves repairing cracks and regularly removing chimney obstructions. Moreover, households must place the pots on the openings correctly, and cover the second pot when it is not in use in order to prevent smoke from escaping.

In addition to providing the stoves, Gram Vikas conducted the standard information campaigns that NGOs run when they introduce a new program for households that have won the lottery. Specifically, during construction they held training sessions on proper use and maintenance (see online Appendix Figure 2 for an example of the training materials). Among households that received a stove in the first wave, almost 70 percent report that they attended a training session. Moreover, Gram Vikas identified individuals in each village who used their stoves correctly and hired (with a small stipend) them to promote proper use and alert Gram Vikas when the stoves were in need of repair. Of those who received a stove in the first wave, 62 percent report knowing who this "promoter" is, 48 percent report that they attended a meeting with the promoter, and another 47 percent state that they received a visit from the promoter to discuss stove use. In total, about 86 percent report either having Gram Vikas or the promoter provide training on the stove (either through a meeting or visit).

B. *Sample, Timeline, and Experimental Design*

In the summer of 2005, Gram Vikas obtained permission from 42 villages to participate in the study. In a decision unrelated to the study, three villages withdrew from all Gram Vikas activity. As a result, we added five additional villages in June 2007. Therefore, a total of 2,575 households in 44 villages participated in the study.

After we completed the baseline survey in each village (in 2006 for the majority of villages, and in 2007 for the additional five villages; see timeline in online Appendix Figure 3), a village meeting was conducted. At each meeting, Gram Vikas explained that the stoves were being built in three waves, and that the households would be randomly assigned to each wave.[8] Next, a public lottery (monitored by the research team) was conducted to choose the first third of households in the village that would be offered a GV improved stove. Gram Vikas completed the first wave of stove construction and user training between September 2006 and March 2007.

After we conducted the midline survey, the second round of village meetings occurred. A lottery was conducted to choose households that would be offered a stove in the second wave of construction. From May 2009 to April 2010, the second round of stove construction and training occurred. Note that during this time, there was also a big push by Gram Vikas to repair or rebuild stoves from the first wave of construction.

II. Data

A. *Data Collection*

Throughout the study, we conducted a series of surveys to create a panel dataset on stove use, smoke exposure, health, stove breakages and repairs, and fuel use. Online Appendix Figure 4 provides a summary of the surveys and their sample sizes, while the online Appendix describes each survey that we conducted in more detail. Here, we provide a summary of the key variables of interest.

We collected comprehensive data on the sociodemographic characteristics of each household. This data includes household composition (size, as well as each member's age, sex, and relationship to the head of household), demographics (education levels, caste, religion), economic indicators (assets, indebtedness), and consumption patterns. In addition, for each household member, we collected measures of productivity, such as employment status, time-use patterns for adults over the last 24 hours, and school enrollment and attendance for children.

Through a series of surveys, we collected information on stove use. This included the types of stoves a household owned, meals cooked with each type of stove over the previous week, repairs and maintenance activities surrounding the stoves, and fuel expenditures (both money and time). In addition, we collected information on beliefs about the efficacy of the stoves (for example, whether they use less fuel) and on satisfaction with the stoves.

To measure smoke exposure, the team measured exhaled carbon monoxide (CO) with a Micro Medical CO monitor.[9] CO is a biomarker of recent exposure to air

[8] It is unlikely that households that refused to participate in the lottery would have benefited so much more from the stoves that the results would have been different if they had been included. First, relatively few households (about 208) refused to participate in the lottery. Second, while these households tend to be richer, they were also less likely to own any type of clean stove at baseline and cooked fewer meals in open areas outside, thus perhaps signaling their lack of demand to prevent exposure to indoor air pollution. Most importantly, the CO in the primary cook's breath at baseline was the same in those households.

[9] Note that we did not measure ambient pollutants (neither CO nor PM). Ambient measures alone are less interesting to measure than exposure, as individuals may undertake fewer behaviors to protect themselves from

pollution from biomass combustion, and therefore it can be used to proxy an individual's personal exposure to smoke from their stoves. Furthermore, it is an inexpensive way to proxy for inhalation of particulate matter, which has been shown to be an important determinant of infant mortality and life expectancy (see, for example, Chay and Greenstone 2003a, 2003b; Chen et al. 2013; Currie and Neidell 2005; Jayachandran 2009; Arceo-Gomez, Hanna, and Oliva 2012).

We collected two types of health data. First, we conducted detailed health recall surveys where we inquired about symptoms (coughs, colds, etc.), infant outcomes, and health expenditures. We complemented these data with physical health checks for biometric measurements, such as height, weight, and arm circumference. During the physical health check, we administered spirometry tests designed to gauge respiratory health by measuring how much air the lungs can hold and how well the respiratory system can move air in and out of them. In contrast to peak flow tests, which are easier to administer, spirometry readings can be used to diagnose obstructive lung disorders (such as chronic obstructive pulmonary disease (COPD) and asthma), and also restrictive lung disorders.[10] Further, this test is the only way to obtain measurements of lung function that are comparable across individuals (Beers and Berkow 1999). The tests were conducted using the equipment directions, as well as guidelines from the American Association for Respiratory Care.[11]

Finally, throughout the study, we compiled Gram Vikas's administrative data. Specifically, we collected data on lottery participation, treatment status, and outcomes.

B. *Sample Statistics*

Table 1 provides information on household-level baseline demographic characteristics and stove usage. For each variable, means are provided in column 1, standard deviations in column 2, and the sample sizes in column 3. As panel A indicates, the households were very poor, with an average monthly per capita household expenditure of about US$12 (rupee 475). Forty-three percent of households belonged to a disadvantaged minority group. A little less than half of households had electricity, which made electric stoves an impractical option. Schooling outcomes

smoke if ambient measures fall and thus could, in fact, end up with a higher level of exposure. If we conducted only ambient measures we could see a decline, even though their actual exposure may not have decreased due to these behavioral changes. We focused on CO, which has been argued to be a good proxy for PM. Collecting data on PM exposure is difficult in this setting: tubes must be attached to the subjects for 24 hours and the equipment requires controlled temperature, careful transferring of samples, and proper laboratories for testing. Given the conditions of rural Orissa, controlling the samples would be near impossible on such a large scale. However, McCracken and Smith (1998) report a strong correlation between the average concentrations of CO and $PM_{2.5}$ in the kitchen during water boiling tests. They conclude that this implies "the usefulness of CO measurements as an inexpensive way of estimating $PM_{2.5}$ concentrations," even if it is not an exact proxy (see Ezzati et al. 2002 for a discussion of this).

[10] According to the Global Initiative for Chronic Obstructive Lung Disease, the results can be used to assess whether participants have COPD. There are two main forms of COPD, chronic bronchitis and emphysema; complications include heart failure, pneumonia, severe weight loss, and malnutrition.

[11] A manual spirometer was used in the baseline, continuous health survey, and a portion of the midline. The enumerators would take up to seven readings for each individual, until there were at least three satisfactory readings and at least two FEV1 readings within 100mL or 5 percent of each other. Electronic spirometers were adopted halfway through the midline. The new machines indicated when satisfactory readings had been completed and saved the best reading for each individual.

Table 1—Baseline Household Demographic Characteristics and Stove Usage

	Mean (1)	Standard deviation (2)	Observations (3)
Panel A. Sociodemographic characteristics			
Household size	6.57	3.49	2,530
Monthly per capita household expenditures	475.28	299.00	2,494
Minority household (scheduled caste or tribe)	0.43		2,516
Has electricity in household	0.47		2,529
Male head ever attended school	0.69		2,086
Male head literate	0.58		2,101
Female head ever attended school	0.32		2,373
Female head literate	0.20		2,378
Female has a savings account	0.67		2,413
Panel B. Stove ownership and use			
Traditional stove	0.99		2,481
Any type of "clean stove"	0.23		2,481
Improved stove	0.01		2,481
Kerosene	0.10		2,481
Biogas	0.03		2,481
LPG	0.04		2,481
Electric	0.11		2,481
Coal	0.00		2,481
Cooked most meals with traditional stove in last week	0.93		2,453
Meals cooked last week	13.75	4.08	2,482
Meals cooked last week with traditional stove	12.61	4.65	2,470
Primary cooks (% female)	0.96		2,483
Panel C. Number of meals cooked each week, by stove location			
Open area	7.46	7.19	2,515
Semi-open area	5.09	6.86	2,515
Enclosed area	0.86	3.18	2,515
Panel D. Fuel			
Ever used wood as fuel	0.99		2,458
Minutes spent gathering wood yesterday (if gathered wood)	313.62	274.99	282
Wood used for last meal (in kg)	4.35	5.60	1,163
Meals per bundle of wood	5.04	7.24	2,074
Household gathers wood	0.83		2,515
Ever bought wood	0.35		2,468
Ever sold wood	0.20		2,101

Notes: This table provides sample statistics on the baseline demographics characteristics and stove usage for households. The top 1 percent of values is dropped from continuous variables.

are discouraging: only 69 percent of the household heads had attended school, and just 58 percent self-reported being able to read. Similarly, only 32 percent of the female household heads (or spouses of one) had attended school, with just 20 percent self-reporting that they were literate.

There was a large dependence on traditional stoves and fuels for cooking (Table 1, panel B). Most households (99 percent) owned at least one traditional cooking stove (see online Appendix Figure 1, panel A). About a quarter of households owned any type of low-polluting stove, primarily electric (11 percent) and kerosene (10 percent). Only 23 out of 2,481 households had an improved stove from a previous program that Gram Vikas had conducted several years earlier. Despite the fact that many households owned a low-polluting stove, most (93 percent) continued to use the traditional stove as their primary stove, with an average of 12.6 meals (or 92 percent of all meals) cooked on one over the last week. Unsurprisingly, 96 percent of those reporting that they primarily cooked household meals were female.

Given that households often had more than one stove, they tended to cook in more than one location (Table 1, panel C). On average, about one meal per week was cooked in an enclosed area, and about five meals per week were cooked in a semi-enclosed area. About 7.5 meals per week were cooked outside. It is noteworthy that open-fire stoves pollute enough that they produce significant exposure even when used outdoors, and rates of COPD appear similar for those who cook inside or outside with open flame stoves (Balakrishnan et al. 2002 and Johnson et al. 2011).

Households relied heavily on wood for fuel (Table 1, panel D), with 99 percent reporting having used wood as fuel at some point. On average, about 4.5kg of wood was used to cook the household's last meal. Fuel was typically obtained by a combination of collection and purchases: 83 percent report having ever gathered wood, and 35 percent report having ever bought wood for cooking. About 20 percent of households also report having ever sold wood. About 10 percent report having collected wood the previous day, spending about five hours, on average, if they did so.

Table 2 presents baseline health statistics. Panel A reports on primary cooks, while panel B reports on children. The primary cooks had high levels of smoke exposure: the average CO reading is 7.77 ppm, where a reading between 6–9 ppm indicates smoke within the lungs and a reading of 10 ppm or more indicates a high level of smoke within the lungs.[12] About 27 percent of them scored a reading of 10 ppm or more, which, following the back-of-the-envelope calculation by Beltramo and Levine (2010), suggests that they had exposure levels that were equivalent to smoking 10 cigarettes per day (note that few women reported that they smoked). In contrast, lung function measurements were in the normal range. We observe a mean FEV1/FVC of about 90, which suggests that, on average, participants did not have COPD.

In general, self-reported illness levels among the primary cooks were high. Almost 90 percent report having had any type of symptoms in the past 30 days. Symptoms that are typically associated with smoke exposure were abundant: about half self-reported having had a cough or cold in the last 30 days, about 49 percent report having had a headache, and about 28 percent report having had sore eyes. In contrast, very few individuals report that they experienced tightness in their chest (4 percent) or wheezing (1 percent). Relatively speaking, health expenditures were high, with the primary cooks reporting that they spent about US$1.50 in the last month.

Children had high levels of CO exposure and poor health outcomes in the baseline (Table 2, panel B). Their CO levels were, on average, 6.48 ppm.[13] This suggests

[12] The baseline CO in exhaled breath is slightly lower than in the RESPIRE study, which found a baseline CO rate of about 9 ppm (Diaz et al. 2007). However, it is similar to the control group mean in the RESPIRE study of about 7 ppm that was observed throughout the course of that study (Smith-Sivertsen et al. 2009).

[13] Note that in the baseline and Continuous Health Survey (CHS), only children approximately 9 years and older were tested for CO exposure, as it is difficult to test younger children. Based on a doctor's assessment and field testing, we lowered the age restriction and collected CO measures for children older than 5 years in the midline and endline.

Table 2—Baseline Carbon Monoxide Exposure and Health

	Mean (1)	Standard deviation (2)	Observations (3)
Panel A. Primary cooks			
Carbon monoxide (CO)	7.77	6.26	2,042
FEV1	1.97	0.37	1,720
FVC	2.30	0.44	1,718
FEV1/FVC × 100	89.64	6.11	1,679
BMI	18.90	2.51	2,167
Cold or cough	0.52		2,511
Phlegm	0.13		2,510
Headache	0.49		2,511
Sore eyes	0.28		2,511
Wheezing	0.01		2,510
Tightness in chest	0.04		2,509
Any illness	0.87		2,511
Health expenditures in the last month	70.05	183.24	2,258
Panel B. Children			
Carbon monoxide (CO)	6.48	5.29	517
BMI	−1.85	1.29	2,700
Cough	0.40		3,343
Consulted health provider about fever	0.27		3,282
Earache	0.09		3,343
Skin irritation	0.13		3,342
Vision problems	0.01		3,343
Hearing problems	0.01		3,343
Vomiting	0.08		3,343
Diarrhea	0.08		3,343
Abdominal pain	0.14		3,342
Worms	0.09		3,339
Weakness	0.22		3,342
Any illness	0.73		3,343
Health expenditures in the last month	46.15	102.11	3,249

Notes: This table provides sample statistics on baseline IAP and health for women, primary cooks, and children. For continuous variables, the top 1 percent of values are dropped. BMI for children is standardized using values from the 2000 US CDC Population of Children.

that they had an average CO level similar to someone who smokes about 7 cigarettes per day. About 20 percent of the children had a reading of 10 ppm or higher, which is equivalent to being a heavy smoker. Children were malnourished, with an average BMI nearly two standard deviations below the norm, according to the 2000 US Centers for Disease Control measurement of the child population (two standard deviations below the norm is generally considered an indicator of stunting). Parents report that 73 percent of the children had some form of illness in the past month. About a quarter of parents consulted a health care provider for a child's fever in the last month, with an average of about US$1 spent on all healthcare costs during this period. Coughs were the most prevalent symptom, with about 40 percent of all children having had one in the last 30 days. Other illnesses that could be associated with indoor air pollution include ear infections (9 percent), skin irritation (13 percent), and vision problems (1 percent).[14]

[14] In online Appendix Table 1, we present the coefficient estimates from regressions of baseline CO and health variables on the number of meals cooked with a low-polluting stove in the last week at baseline, conditional on village fixed effects. For primary cooks, more meals cooked with any type of low-polluting stove was associated with

III. Empirical Framework and Experimental Validity

A. *Empirical Framework*

Most of the evidence on health improvements from reductions in indoor air pollution is based on the association between clean stove usage and health in observational data. However, those who choose to use a clean stove may generally value health more than those who do not and thus may also undertake other health investments, either of which would lead to better health. In this case, our estimated coefficients would be biased upwards. Alternatively, the improved stoves may be disproportionately used by the sick, which would cause the estimated relationships to be biased downwards.

The experimental design we propose allows us to solve these endogeneity problems by comparing winners and losers from the stove lottery. We begin by estimating the reduced form effect of winning the stove on a series of outcomes, including stove use, CO exposure, health, and other non-health stove outcomes (such as fuel use and cooking time). Specifically, we estimate:

$$Y_{ihvt} = \beta_0 + \beta_1 T_{ihvt} + (\boldsymbol{\delta}_v \times \gamma_t) + \varepsilon_{ihvt}, \tag{1}$$

where Y_{ihvt} is the outcome of interest for individual i in household h in village v at time t. T_{ihvt} is an indicator variable that equals 1 if the household was in the treatment group at time t. As we stratified the sample by village during the randomization, and treatment and control households were surveyed at about the same time within each village, we include village × survey month-year fixed effects $(\boldsymbol{\delta}_v \times \gamma_t)$; i.e., there are separate fixed effects for all observations from a village in a given month-year (e.g., January 2010). For CO exposure, health, and non-health stove outcomes (when possible) we additionally include the baseline value of the outcome to gain additional precision. β_1 is our key parameter of interest; the random assignment of T_{ihvt} ensures that β_1 will be an unbiased estimate of the effect of being offered a stove.

To fully exploit the four years of follow-up, we additionally estimate how the treatment effect varies over time. The effect of being offered a stove may change throughout time for a variety of reasons. The effect may decline over time if the stoves break or fall into disrepair, proper use declines, or if individuals feel healthier and compensate with other unhealthy behavior (i.e., smoking). Alternatively, the effect may increase if households learn how to use the stoves better or use them more as they learn about the benefits of the stoves over time. To capture this change, we interact the treatment effect (T_{ihvt}) with a set of indicator variables (I_k) for whether the observation falls within a given year after stove distribution $k = \{1, 2, 3, 4\}$:

$$Y_{ihvt} = \beta_0 + \textstyle\sum_{k=1}^{4}\big(\beta_k(T_{ihvt} \times I_k)\big) + (\boldsymbol{\delta}_v \times \gamma_t) + \varepsilon_{ihvt}. \tag{2}$$

better lung functioning (column 3), higher BMI (column 4), and a lower likelihood of sore eyes (column 6) and wheezing (column 9). For the children, there is a significantly negative correlation between each additional meal cooked with a clean stove in the last week and smoke exposure. However, we do not observe a relationship between meals and health. The signs of the coefficients suggest improved health outcomes, but they would not be judged to be statistically significant by conventional criteria, despite the relatively large sample sizes.

In equation (2) there are now four parameters of interest (β_1, β_2, β_3, and β_4), which capture the effect of having won the lottery within one year of the stove being built, within 13 to 24 months of the stove being built, etc. Due to the timing of Lottery 2 and the surveys, β_1 is identified from winners of both Lottery 1 and 2, but the other βs are only identified from the Lottery 1 winners.

To scale the results, we also estimate the effect of using any type of low-polluting stove on CO exposure using an instrumental variables strategy. We estimate

$$Y_{ihvt} = \theta_0 + \theta_1 Use_{ihvt} + (\boldsymbol{\delta}_v \times \boldsymbol{\gamma}_t) + \varepsilon_{ihvt}, \tag{3}$$

where Use_{ihvt} is either a measure of whether the household owns a low-polluting stove or the number of meals cooked with a good condition, low-polluting stove over the last week. As selected individuals may choose whether to take up a stove, an OLS estimate of θ_1 would be biased. Thus, we use the treatment variable (T_{ihvt}) as an instrument for Use_{ihvt} in equation (3).

Finally, note the following specification details. First, the household-level equations are weighted to account for household splits and mergers. Second, for all regression analysis, the standard errors are clustered at the household level, which is the unit at which the treatment was assigned.

B. *Verification of Experimental Validity*

There are two primary threats to the empirical design. First, the randomization may have produced imbalanced groups either by chance or if the randomization process was somehow corrupted. It is unlikely that the process was corrupted, as the lotteries were publicly conducted and our research team monitored each of them. Nonetheless, in online Appendix Table 2A and 2B, we provide results from a test of the randomization across baseline demographics, stove use, and health for the primary cooks and children across Lottery 1 winners, Lottery 2 winners, and those who lost both lotteries. The groups are well-balanced across the 59 baseline characteristics that we consider, with only 10 percent of the differences across groups significant at the 10 percent level or more (as predicted by chance). Further details are described in the online Appendix.

Second, poor areas are often characterized by seasonal migration. Moreover, individuals may not have been home when our enumerators visited them if they were working in the fields, etc. Attrition would be most problematic if it is correlated with treatment status (e.g., households that obtained a new stove were less likely to migrate). We tried to minimize overall attrition by revisiting households that we could not initially locate, as well as conducting the surveys in the evening when individuals were likely to be at home. As online Appendix Figure 4 shows, we find about 94 percent of the households in the first main two surveys and about 81 percent in the last survey. Results in online Appendix Table 3 (details are provided in the online Appendix) fail to produce any meaningful evidence of differential attrition across the treatment and control groups, implying that differential attrition is not a source of bias in the subsequent regressions.

IV. Results

This section is broken into four segments. We begin by examining the relationship between stove ownership and usage (subsection IVA). Next, we explore the relationship between treatment status and CO exposure (subsection IVB) and health (subsection IVC). Finally, we explore the relationship between the stoves and household expenditures on fuel and repairs (subsection IVD).

A. *Improved Stove Ownership and Use*

In the real world, the health gains of a stove will be achieved if households choose to take up a stove, keep the stove in good condition, use it properly, and use it regularly (see Figure 1 for the causal chain). Moreover, the initial treatment effect may change over time, as stoves may deteriorate and/or individuals may update their beliefs about how to use the stoves or their expected benefits.

We begin the analysis by exploring the effect of treatment status (i.e., being offered a GV improved stove) on stove ownership and use over time. Figure 2 plots the β_k's and their 95 percent confidence intervals from a specification where we interact the treatment status indicator with indicators for months since stove construction in the village in six-month intervals, after adjustment for the village $\times$ survey month-year fixed effects (this is a modified version of equation (2)). As shown in panel A, over 70 percent of households that won Lottery 1 built a GV stove during the first six months of the program. Lottery 2 winners did not look very different than Lottery 1 winners.[15] The fraction of households with installed GV stoves, regardless of their condition, declined over time. In the final year, the rate of stove ownership increased again as GV repaired broken stoves from Lottery 1 during the construction of stoves for Lottery 2 winners. Appendix Table 4 illustrates that the reasons for not building/rebuilding a stove changed over time as households learned about the stoves: the fraction of households who claimed they were not interested in the stove increased from 7 percent in Year 1 to about 26 percent in Year 4. Further, the fraction of households that destroyed their stove, presumably to create space in their homes, increased from 2 percent in Year 1 to 32 percent by Year 4.

Next, we explore how the offer of the GV improved stove changed the number of meals cooked with any type of low-polluting stove in good condition over the previous week. This outcome captures the intensity of use and is the most direct measure of the improved stoves' potential impact on health. Figure 2, panel B shows that treatment households cook about three more meals a week than the control households on a good condition, low-polluting stove during the first year.[16] The effect

[15] The reasons for not taking up a stove varied, as shown by online Appendix Table 4. In the first year, about 28 percent who chose not to take a stove did so because the stoves were inconvenient: either they did not believe that they had sufficient kitchen space or the fact that the stove was not the right fit for their family size. Only 6 percent claimed that they were not building it because they had a better stove. About another quarter claimed that they were planning on building a stove soon.

[16] Looking at improved stoves in Ghana, Burwen and Levine (2011) also find that individuals do not completely reduce their use of the traditional stoves when given an improved stove, with the treatment group using an average of 1.4 traditional stoves as compared to 1.9 in the control group. In fact, they returned to three of eight villages about

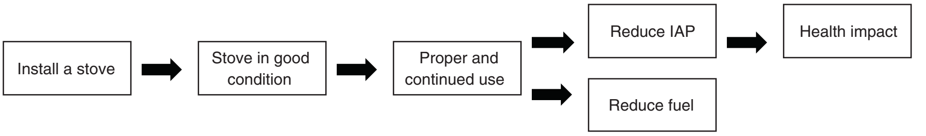

Figure 1. Causal Chain

Note: This figure traces out the behavioral chain necessary to observe health and fuel impacts after a stove offer is made.

falls over time and picks up in the fourth year, when there was a big push by Gram Vikas to construct and retrain households during the Lottery 2 construction.

Table 3 formally tests the effect of the stove offer on take-up and use. Panel A provides estimates of the overall treatment effect (equation (1)) and panel B provides estimates of the overall treatment effect by years since stove construction in the village (equation (2)). Since the stove use behavior of Lottery 1 and Lottery 2 winners was not significantly different, we group them together. We include the F-statistics and p-values to compare year 1 against each subsequent year at the bottom of the table, as well as the control group means.

The estimates illustrate that take-up was far from universal and proper usage was substantially smaller than take-up. About 6 percent of the control group took up the GV improved stoves, and the treatment group was 62 percentage points more likely to have a GV stove than the control group (Table 3, column 1).[17] By Year 3, the treatment group was only 44 percentage points more likely to have a GV stove (the p-value of the difference between Year 1 and Year 3 was 0.000); the figure increases again in Year 4 as a result of the rebuilding campaign. Considering all low-polluting stoves, the overall treatment effect falls to about 47 percentage points, as about a quarter of the control group had a low-polluting stove of any type (Table 3, column 2).

The stove condition may deteriorate over time due to normal wear and tear coupled with insufficient maintenance. This deterioration could lead to increased levels of smoke within the household. As column 3 in Table 3 reports, many households did not undertake the investments necessary to keep the stoves in good condition: the treatment effect on the proportion of GV improved stoves in good condition is 36 percentage points over the entire period. The effect again is high in the first two years, falls in Year 3, and increases again during the big push in Year 4 (Table 3 column 3, panel B).[18]

On net, households did not use the stoves regularly. On average, treatment households cooked about three more meals per week on any good condition, improved stove; out of a total of 14 cooked meals per week, this is approximately 20 percent

eight months after the stove installation and found that only about half of the improved stoves remained in regular use (i.e., warm to touch or contained reasonable amounts of ash).

[17] The overall take-up rate is not inconsistent with other preventive health products that have demonstrated health effects (see Dupas 2011 for a discussion).

[18] Online Appendix Figure 5 helps to explain this finding by showing that the percent of Lottery 1 household winners that report ever having had a crack in the stove was 74 percent and the comparable figure for Lottery 2 is 67 percent, which is striking since they were followed for only one year.

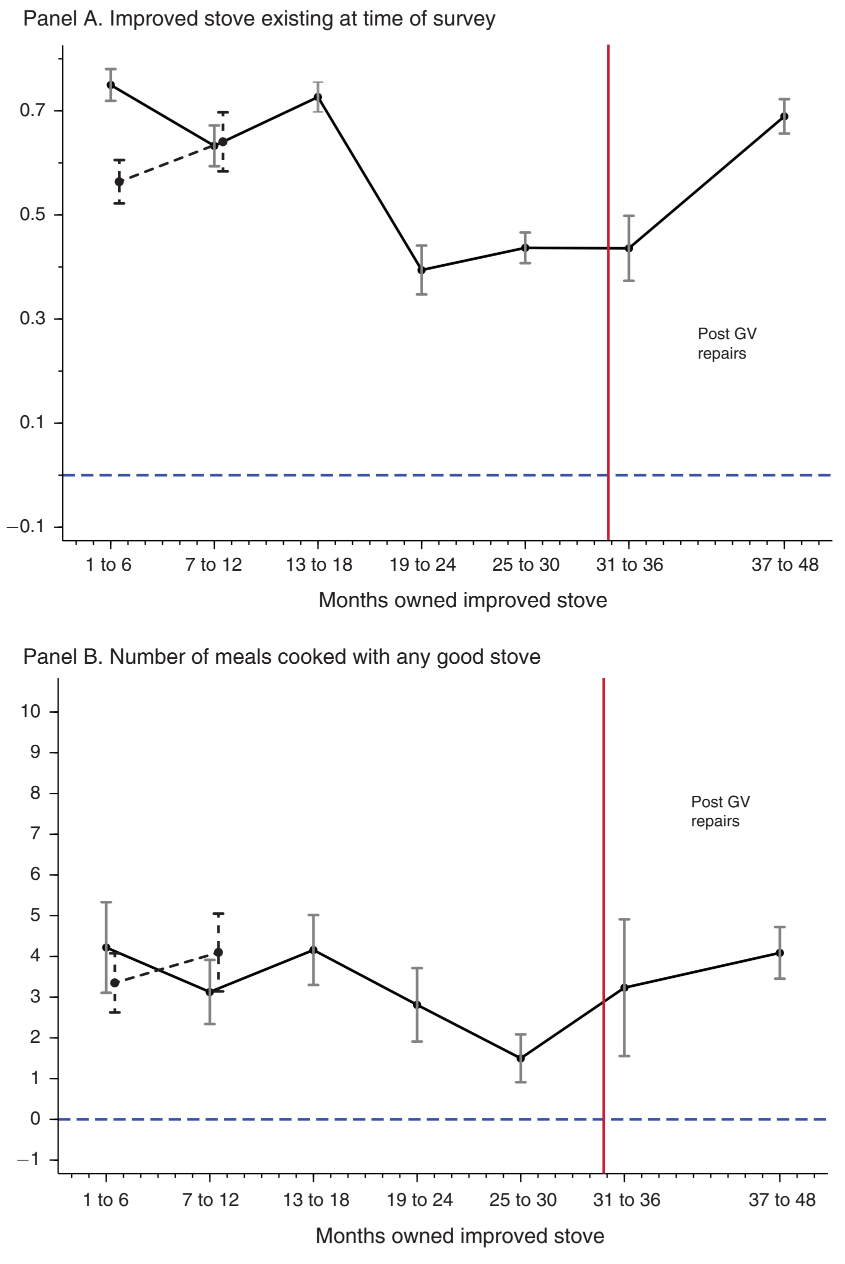

Figure 2. Stove Ownership and Usage, by Time

Notes: These figures show the difference in stove usage between the treatment and control groups, by months since stove construction in the village and lottery status, conditional on village × month of survey × year fixed effects. Regressions are weighted to account for splits and mergers. The solid line signifies Lottery 1, while the dashed line signifies Lottery 2. The bars represent the ninety-fifth percent confidence interval.

TABLE 3—REDUCED FORM EFFECT OF STOVE OFFER ON TAKE-UP AND USAGE

	Gram Vikas improved stove at time of survey (1)	Any low-polluting stove (2)	Gram Vikas improved stove in good condition (3)	Number of meals cooked with any good condition, low-polluting stove (4)	More than 75% of meals on a good stove (5)
Panel A. Overall treatment effect					
Treat	0.618***	0.469***	0.364***	3.124***	0.195***
	(0.011)	(0.012)	(0.011)	(0.177)	(0.011)
Panel B. By months since stove construction					
Treat × I(0 to 12 mo)	0.654***	0.478***	0.364***	3.472***	0.219***
	(0.012)	(0.014)	(0.012)	(0.229)	(0.016)
Treat × I(13 to 24 mo)	0.670***	0.500***	0.430***	3.618***	0.243***
	(0.014)	(0.018)	(0.017)	(0.327)	(0.023)
Treat × I(25 to 36 mo)	0.441***	0.396***	0.286***	1.758***	0.096***
	(0.014)	(0.015)	(0.015)	(0.287)	(0.016)
Treat × I(37 to 48 mo)	0.722***	0.516***	0.430***	4.007***	0.241***
	(0.019)	(0.018)	(0.020)	(0.326)	(0.021)
Observations	18,967	17,459	15,371	6,792	6,792
F-stat Yr1 = Yr2	1.324	2.033	15.57	0.139	0.764
Prob > F Yr1 = Yr2	0.250	0.154	0.000	0.710	0.382
F-stat Yr1 = Yr3	237.7	26.71	23.18	26.68	32.54
Prob > F Yr1 = Yr3	0.000	0.000	0.000	0.000	0.000
F-stat Yr1 = Yr4	16.74	5.576	11.15	2.744	0.960
Prob > F Yr1 = Yr4	0.000	0.018	0.001	0.098	0.327
Control group mean	0.0643	0.245	0.0435	2.321	0.0817

Notes: This table provides information on stove ownership and usage over time. All regressions are estimated using OLS, include village × month of survey × year of survey fixed effects, and standard errors are clustered at the household level. Regressions are weighted to account for household splits and mergers. Good condition is defined as those stoves reported to be in good condition as observed by the enumerator.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

more (Table 3, column 4). This fell from 3.472 in Year 1 to 1.758 by Year 3.[19] This average masks some considerable heterogeneity. Specifically, we observe a 20 percentage point increase in households using a good stove for more than 75 percent of meals in the previous week (Table 3, column 5); this falls to 9.6 percentage points by Year 3 and increases again after the stove reconstruction in Year 4. As Figure 3 shows, use of the stoves tends to be bimodal, with some treatment households never using their good condition GV stove and others using it for essentially all of their meals.

[19] As shown in online Appendix Table 5, we observe some heterogeneity in treatment effect. For example, those who had an improved stove at the time of baseline were less likely to take-up a GV stove and also cooked fewer meals with *any* type of improved stove in the follow-ups. We do not observe any heterogeneity based on per capita consumption. However, we observe that households with a less educated head and those from a disadvantaged minority group more likely to take up an improved stove, consistent with the fact that these types of groups were less likely to have any form of improved stove at baseline.

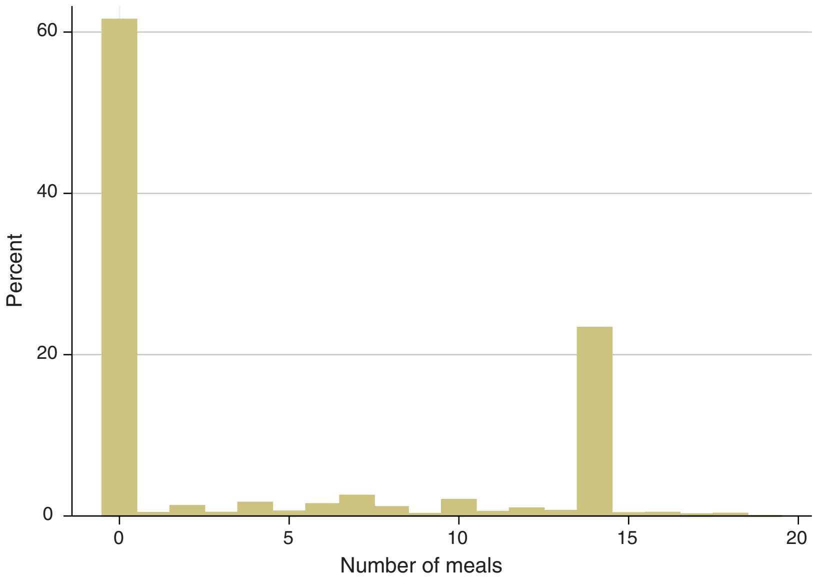

Figure 3. Number of Meals Cooked for Treatment Households Who Own a Good Condition GV Stove

Note: This figure explores the number of meals cooked with a GV stove in the last week for treatment households with a GV stove ever built.

If households do not use the stoves correctly, each additional meal cooked will not reduce smoke inhalation to the fullest possible extent. For example, a failure to cover the second pot opening when it is not in use will allow smoke to enter the kitchen through this opening. Similarly, if households fail to clean the chimney regularly, it will become blocked and smoke will enter the kitchen when the improved stove is used.[20] It is difficult to measure proper use. Often, use is gauged through controlled kitchen tests, but households may use the stove correctly when they are being observed by researchers even if they do not typically use it properly. Similarly, self-reported measures of use may be biased upwards if households feel judged by the enumerators.[21]

Nonetheless, we collected self-reported measures of proper use. As Figure 4 shows, for the sample of those who own a stove in good condition, only about 60 percent report that they use the stoves properly, where proper use is defined as cleaning the stove in the last week, using the stove in the last week, not elevating the cook pot during use, and using the two pots correctly.

[20] A self-reported good use does not necessarily mean that the stove will be in good condition: Dutta et al. (2007) find that even when households self-reported regular cleaning by dropping sand bags from the top of the chimney, the chimneys could often became clogged four to five months after installation if the cleaning was not done properly.

[21] Another reason that smoke inhalation may not be reduced to the fullest extent possible is if the stoves induce individuals to cook inside and the smoke exposure from a clean stove inside is worse than the smoke exposure from a traditional stove outside. However, we find no evidence that treatment households increased the number of meals cooked indoors. Moreover, we do not observe any heterogeneity in the treatment effect by the number of meals the household cooked outside at the baseline.

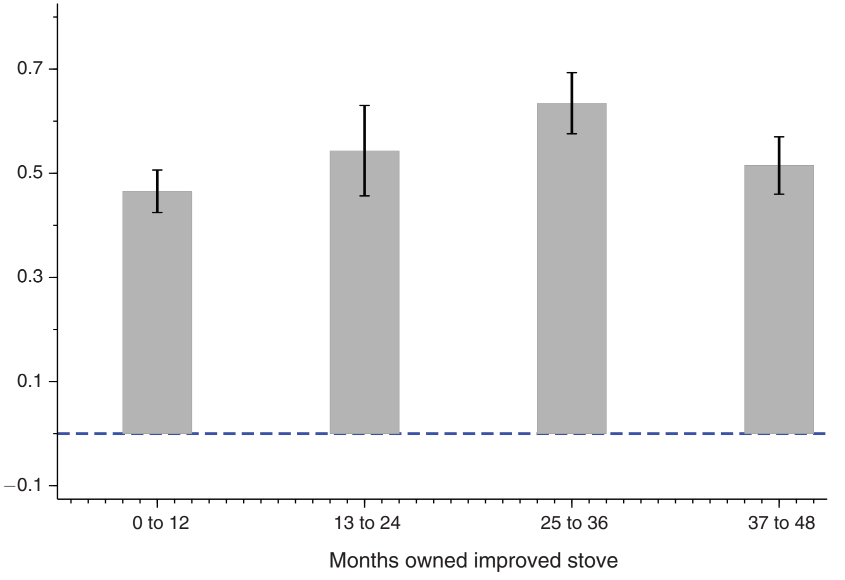

Figure 4. Proper Use for Those Who Owned a Good Condition GV Stove

Notes: Good condition is defined as observed by the enumerator. Proper use is defined as cleaning the stove in the last week, using the stove in the last week, not elevating the cookpot during use, and using the two pots correctly.

In summary, we find that stove behavior and use in real-world settings differs considerably from controlled laboratory tests. Take-up of GV stoves was only about 60 percentage points higher in treatment households than control ones, despite the fact that the stoves were highly subsidized. The share of households that maintained an improved stove in good condition was substantially smaller at 36 percentage points, and out of these, 40 percent self-reported that they did not properly clean and use the stoves to minimize indoor air pollution. In practice, treatment households also continued to use their traditional stoves, cooking only about three extra meals per week on any type of low-polluting stove in good condition.

B. *Effects on Smoke Inhalation*

We now test whether being offered a stove caused changes in smoke inhalation. Following Diaz et al. (2007), we measured Carbon Monoxide (CO) in exhaled breath to measure smoke inhalation. As discussed in the data section, CO is a biomarker of recent exposure to air pollution from biomass combustion, and therefore it can be used to proxy an individual's personal exposure to smoke from cooking stoves.

Table 4 provides a reduced form analysis of the effect of stoves on smoke exposure for those who identified themselves as primary cooks in the baseline and for children who were old enough to be tested.[22] All specifications include the baseline

[22] As Pitt, Rosenzweig, and Hassan (2010) discuss, indoor air pollution is unlikely to be evenly distributed within the household, with the highest incidence likely borne by those who do most of the cooking; thus, we test the effect on primary cooks in the household.

Table 4—Reduced Form Effect of Stove Offer on Carbon Monoxide Exposure

	Primary cooks (1)	Children (2)
Panel A. Overall treatment effect		
Treat	−0.230	−0.120
	(0.196)	(0.181)
Panel B. By months since stove construction		
Treat × I(0 to 12 mo)	−0.534*	−0.318
	(0.281)	(0.288)
Treat × I(13 to 24 mo)	−0.173	−0.107
	(0.489)	(0.445)
Treat × I(25 to 36 mo)	0.072	−0.152
	(0.317)	(0.213)
Treat × I(37 to 48 mo)	0.099	0.326
	(0.434)	(0.413)
Observations	4,234	4,401
F-stat Yr1=Yr2	0.399	0.161
Prob > F Yr1 = Yr2	0.528	0.688
F-stat Yr1 = Yr3	2.159	0.227
Prob > F Yr1 = Yr3	0.142	0.634
F-stat Yr1 = Yr4	2.235	2.771
Prob > F Yr1 = Yr4	0.135	0.096
Control group mean	7.128	5.460

Notes: This table provides the reduced form effect of being offered a GV stove on carbon monoxide levels. All regressions are estimated using OLS, include village × month of survey × year of survey fixed effects, include baseline carbon monoxide, and standard errors are clustered at the household level. The top 1 percent of values are dropped. Primary cook is defined as the individual who reported, in the baseline survey, cooking the majority of meals in the household during the last week.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

values of the outcome variable and village × survey month-year fixed effects, and the standard errors are clustered at the household level.[23]

On average, we observe limited effects on the CO concentrations in respondents' breath, with the effect decreasing over time. On net, CO levels fall slightly for primary cooks, but the effect is not significant (Table 4, panel A, column 1). In terms of magnitude, it is small, being 3.1 percent of the mean and 3.6 percent of a standard deviation. We observe a meaningful reduction in primary cook's CO breath concentrations during the first year. Specifically, we find a 0.53 ppm reduction (7.5 percent of the control group's mean) in the CO concentration during Year 1 for primary cooks relative to the control group, when stove usage is at its highest. Thus, to the extent that they were used in the first year, they were effective in reducing CO, which supports ARTI's laboratory results that these stoves can be effective at reducing exposure to indoor air pollution. However, as usage declined so did the effect

[23] If the baseline value is missing, we assign the average of the baseline variable. We additionally include an indicator variable that equals one when the baseline value for an individual was imputed.

on CO: the treatment effect for primary cooks falls to −0.173 ppm by the second year of stove ownership and is no longer significant, and is positive and insignificant by Years 3 and 4. The *p*-value of the difference between Year 1 and Year 3 is 0.142, while it is 0.135 between Year 1 and 4 (note if we group Year 3 and 4 together for power, the *p*-value of the difference between Year 1 (and Year 3 and 4) is 0.06). Thus, taken together, the evidence suggests that the effect for primary cooks is positive at the start, but declines over time.[24]

For children, the overall effect is also negative, but statistically insignificant (Table 4, column 2, panel A). While smoke exposure generally decreased in children in the first two years, this effect is not statistically significant at conventional levels (panel B).

To interpret these results, Table 5 reports the results from estimating the effect of owning any type of low-polluting stove on CO exposure with the instrumental variables approach outlined in equation (3). Additionally, we estimate the effect of an additional meal cooked on a good condition, low-polluting stove on CO exposure—as well as the effect of cooking 75 percent of meals with a good condition, low polluting stove on CO—with the same instrumental variables approach. Columns 1–3 estimate the effect for the primary cooks, while Columns 4–6 estimate the effect for children. Note that all specifications include baseline values of the outcome variable and village × survey month-year fixed effects and are clustered at the household level.

Before turning to the results, it is worth noting that the instrumental variable estimates are not equal to the ratios of the relevant reduced form relationships in Tables 3 and 4. This is because the household-level data on the presence of a stove and meals cooked with a low-polluting stove in good condition were collected in a different survey than the individual-level data on the CO breath concentrations. Consequently, we have smaller sample sizes in Table 5 than in Table 4, and the first stage is slightly different.

On average, owning at least one of any low-polluting stoves reduces CO levels by −0.656 ppm for primary cooks and −0.346 ppm for children. These scaled estimates suggest declines of 9.12 percent and 6.67 percent, respectively, in smoke exposure from owning an improved stove, but none of them are statistically different from zero (Table 5, panel A). Owning an improved cooking stove in the first year reduces CO exposure for primary cooks by −1.001 ppm, or 14.0 percent, relative to the control group (Table 5, panel B). By Years 3 and 4, the effect becomes positive and remains statistically indistinguishable from zero.

A comparison of these IV results with the RESPIRE studies' estimates helps to underscore the fundamental differences in approach and meaning of the studies' results. With weekly maintenance and instruction on proper use, as well as the use of stoves for most meals, the RESPIRE intervention produced a reduction in CO exposure of about 60 percent for women and 50 percent for children. These effects are much larger, for example, than the statistically insignificant 6.67 percent reduction in CO concentrations for children that arise from stove ownership within our

[24] As shown in online Appendix Table 5, the effect on CO is larger for less educated households, i.e., households that were more likely to take up a GV stove.

Table 5—IV Effect of Stove Usage on Carbon Monoxide

	Primary cooks			Children		
	Any type of low-polluting stove (1)	Meals on good condition, low-polluting stove (2)	75% of meals on good condition, low-polluting stove (3)	Any type of low-polluting stove (4)	Meals on good condition, low-polluting stove (5)	75% of meals on good condition, low-polluting stove (6)
Panel A. Overall treatment effect						
Stove variable	−0.656	−0.054	−0.826	−0.346	−0.070	−1.126
	(0.506)	(0.065)	(0.998)	(0.546)	(0.072)	(1.164)
Panel B. By months since stove construction						
Stove variable × I(0 to 12 mo)	−1.001*	−0.100	−1.667	−0.201	−0.064	−1.335
	(0.585)	(0.071)	(1.096)	(0.729)	(0.087)	(1.478)
Stove variable × I(13 to 24 mo)	−0.581	0.015	0.226	−0.470	−0.147	−1.946
	(1.305)	(0.172)	(2.651)	(1.189)	(0.174)	(2.295)
Stove variable × I(25 to 36 mo)	0.236	0.075	1.258	−0.792	−0.101	−1.802
	(1.210)	(0.200)	(3.348)	(0.841)	(0.125)	(2.237)
Stove variable × I(37 to 48 mo)	0.394	−0.000	0.215	1.154	0.069	1.286
	(0.983)	(0.092)	(1.581)	(1.182)	(0.112)	(1.860)
Observations	4,117	3,934	3,934	4,268	4,070	4,070

Notes: This table provides the coefficient estimate of the effect of stove usage on carbon monoxide levels, where stove usage is instrumented by treatment status. All regressions include village × month of survey × year of survey fixed effects, include baseline carbon monoxide, and standard errors are clustered at the household level. The top 1 percent of values are dropped.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

study. The fact that the stoves were not used for all meals in our setting is likely responsible for the differences in CO. If households had cooked all meals with an improved stove, there would have been an estimated 18 percent reduction in CO concentrations for children.[25] As we emphasized above, while we cannot be certain that the laboratory effect of our study's stoves are exactly equal to the effect of the RESPIRE study's stoves, it seems safe to conclude that the deterioration of the stoves over time, coupled with improper use (e.g., not covering the second pot), may be responsible for the differences in observed levels of smoke exposure reduction.

There are two other possible explanations for the small estimates in Table 5. First, it is possible that that the GV stoves cause individuals to feel healthier, which leads them to choose activities (like increased cigarette smoking) that would expose them to smoke from other sources. In this case, the impacts on smoke inhalation would be partially or even completely undone by individuals' compensatory responses. However, very few primary cooks (0.7 percent) report smoking during the course of the study, and therefore, changes to such rates appear unlikely to affect the overall CO results. It is also possible that men may smoke more in the household, inducing

[25] This calculation assumed that the effect would be linear. In the IV regressions, each meal results in a −0.07 reduction in CO for children. If all meals were cooked on a good improved stove, we would see a 14 meals × −0.07 = 0.98 reduction in exposure, or about 18 percent of the control group mean.

higher rates of secondhand smoke to women. Online Appendix Table 6 shows the reduced form effect of the treatment on the male propensity to smoke and finds no overall difference (panel A) and no change over time (panel B). Thus, the stoves do not appear to induce compensatory behavior that undoes their beneficial impacts.

Second, the effect of CO concentrations of exhaled breath could be mitigated by two forms of spillovers from the treatment to the control group. Online Appendix Table 7 explores these forms of spillovers. First, treatment households could conduct all the cooking for the control group since they own the improved stove. The data are inconsistent with this possibility, as the total number of meals cooked by treatment and control households was not significantly different during the experiment (the magnitude of the coefficient estimate is near zero, and in fact, negative). Moreover, the number of people whom the treatment household cooked for was not significantly different than that of the control households during these meals. Second, the experiment may cause control households to learn about the dangers of indoor air pollution, which leads them to change their cooking habits to protect themselves from smoke. Using data from our midline survey, we find no difference in the minutes spent cooking at arm's length from one's cooking stove, suggesting that control households were not differentially trying to protect themselves from the smoke.

Conversely, the chimneys could have negatively affected the control group households by increasing *outdoor* air pollution. This would lead to an overestimate of the treatment effects. We do not have much data to refute this possibility, but this would mean that our low effects are upper bound.

C. *Health Outcomes*

This subsection reports the impact of the treatment on a wide range of health outcomes. The results, thus far, suggest that sustainable health effects are unlikely to operate through the channel of reduced smoke inhalation, as there are no sustainable effects on measured smoke inhalation over time. Nevertheless, it is possible that there are unobserved household compensatory responses to the stoves that loosen budget constraints in a way that directly improves health.

Table 6A shows the reduced form effect of the treatment on the health of primary cooks (panel A), children (panel B), and infants (panel C). Note that we report the effects on the additional health measures that we collected in online Appendix Table 8. We first explore respiratory function as measured by spirometer (FEV1 and FEV1/FVC $\times$ 100 in Table 6A, columns 1 and 2, panel A); note that larger spirometry readings indicate greater lung functioning. We find no effect of the treatment on these measures. Turning to the self-reported health measures for both primary cooks and children, we also find no overall effects. In fact, out of the 30 health estimates in both Table 6A and online Appendix Table 8, only 3 are significant at the 10 percent level, which is what is expected by chance. All three of the statistically significant effects have counterintuitive signs, suggesting that the stove offer causes worse health, further underscoring that treatment status appears unrelated to health.

In the presence of so many outcome variables, it can be informative to summarize the results by estimating an average treatment effect across the multiple outcomes. To do this, we standardized all of the health variables to have a mean of zero and

Table 6A—Reduced Form Effect of the Program on Health

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Primary cooks						
	FEV1	FEV1/FVC × 100	Cough or cold	Any illness	Health expenditures	
Treat	0.003	−0.005	0.004	−0.018	2.847	
	(0.018)	(0.003)	(0.015)	(0.014)	(4.138)	
Observations	3,104	3,070	5,324	5,323	5,282	
Control group mean	1.922	0.859	0.410	0.753	45.70	
Panel B. Children aged 13 and under in the baseline						
	BMI	Cough	Consult for fever	Any illness	Health expenditures	Days of school missed last week
Treat	−0.086**	0.012	0.004	0.008	0.701	−0.018
	(0.044)	(0.011)	(0.010)	(0.012)	(1.802)	(0.027)
Observations	6,947	10,500	10,037	10,868	9,765	3,552
Control group mean	−1.415	0.273	0.201	0.573	26.56	0.0974
Panel C. Pregnancy and infant outcomes						
	Birthweight	Infant mortality	Stillbirths and miscarriages			
Treat	36.980	0.011*	−0.004			
	(106.393)	(0.006)	(0.014)			
Observations	630	1,635	1,109			
Control group mean	2,964	0.378	0.0599			

Notes: This table provides the reduced form effect of being offered a GV stove on health. All regressions in panels A–C are estimated using OLS, include village × month of survey × year of survey fixed effects, and standard errors are clustered at the household level. In panel C, the mortality regressions in columns 2 and 3 include village × survey quarter × survey year fixed effects, and the birthweight regression includes village × birth quarter × birth year fixed effects. For all variables except days of school missed last week, we additionally include the baseline value. For continuous variables, the top 1 percent of values are dropped. BMI for children is standardized using values from the 2000 US CDC Population of Children.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

standard deviation of 1, took the average across all outcomes for each observation, and then estimated the effect of treatment status.[26] The results are presented in Table 6B. Not only are none of the estimated effects significant, they are practically very small (panel A). For example, the treatment results are a −0.008 standard deviation change in health across all variables for primary cooks. Furthermore, none of the effects significantly change over time (panel B).

D. *Monetary and Time Costs of Improved Stoves and Self-Reported Satisfaction*

Table 7 examines whether the treatment status causes changes in the monetary and time costs of using and maintaining a household's stoves. Improved stoves can

[26] An observation may comprise a different number of variables due to missing data or due to the fact that surveys may have been conducted at different times. Therefore, we weight each observation by the number of variables that contribute to the average. The results from unweighted regressions are qualitatively the same.

Table 6B—The Reduced Form Effect of Stoves on a Standardized Index of Health

	Primary cooks (1)	Children (2)
Panel A. Overall treatment effect		
Treat	−0.008	0.006
	(0.014)	(0.011)
Panel B. By months since stove construction		
Stove variable × I(0 to 12 mo)	−0.011	0.008
	(0.019)	(0.015)
Stove variable × I(13 to 24 mo)	0.016	0.041
	(0.029)	(0.027)
Stove variable × I(25 to 36 mo)	−0.017	−0.016
	(0.023)	(0.019)
Stove variable × I(37 to 48 mo)	−0.011	0.015
	(0.024)	(0.019)
Observations	7,310	14,188
Control group mean	−0.012	25.86

Note: This table provides the reduced form effect of the stove offer on the standardized indices of health.

Table 7—Time and Cost of Operating Stoves

	Total wood used at last meal (kg) (1)	Total fuel costs last 30 days (rupees) (2)	Time spent cooking last evening meal (minutes) (3)	Number of repairs made in the last year (4)	Time spent on repairs in the last year (minutes) (5)
Panel A. Overall treatment effect					
Treat	0.046	9.482	3.410	2.431***	29.387**
	(0.123)	(6.697)	(3.315)	(0.506)	(14.665)
Panel B. By months since stove construction					
Treat × I(0 to 12 mo)	−0.022	11.800	0.554	6.466***	61.121**
	(0.179)	(11.526)	(5.681)	(1.265)	(26.553)
Treat × I(13 to 24 mo)	0.156	−9.765	12.583*	2.153***	24.766
	(0.168)	(14.421)	(6.603)	(0.630)	(28.437)
Treat × I(25 to 36 mo)	0.235	10.538	1.358	−0.175	−22.865
	(0.249)	(10.851)	(4.524)	(0.358)	(18.578)
Treat × I(37 to 48 mo)	−0.180	18.964	3.800	2.647**	79.574**
	(0.285)	(12.075)	(6.079)	(1.140)	(32.507)
Observations	5,619	4,570	4,652	3,786	3,794
Control group mean	3.373	333	163.9	5.430	202.3

Notes: This table provides the reduced form effect of being offered a GV stove on stove expenditures. All regressions are estimated using OLS, include village × month of survey × year of survey fixed effects, and standard errors are clustered at the household level. The specifications in columns 1–3 also include baseline values. The top 1 percent of values are dropped for continuous variables. Wood use is in kilograms, time variables are in minutes, and fuel costs are in rupees. Regressions are weighted to account for household splits and merges.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

affect expenditures on a number of levels. First, when properly used in controlled conditions, the ARTI stoves require less wood and households received training from Gram Vikas on how to achieve these fuel reductions. As such, the stoves may reduce energy use and hence fuel costs. Second, if the stoves are more efficient both in terms of heating up quickly (e.g., time required to boil water) and the two-pot functionality, cooking times may be reduced. Finally, the new stoves may alter the time spent making repairs. As recognized by the Global Alliance for Clean Cookstoves (2010), these factors are important for adoption (particularly if households are asked to pay for the stoves). Moreover, if the stoves reduce energy use, carbon credits could be used to finance them; this is one of the avenues currently being explored to make them more widely available.

On average, households seem to have been convinced that they should use less wood in the new stoves: more than 60 percent of households report that they believe that GV stoves use less wood (online Appendix Figure 6). However, looking at *actual* use in Table 7, wood use appears unchanged (column 1), while reported total fuel expenditures increases, but not significantly so (column 2). The discrepancy between the laboratory test and the actual expenditures by the households may be due to improper use, or the fact that households now use both the traditional and the improved stove, perhaps simultaneously. Burwen and Levine (2011) observe a similar effect for the type of stove that they evaluate in Ghana. After eight weeks, households took less time and fuel to cook a meal in a carefully controlled test, but there was no significant decline in the actual fuel used by the family.[27] These results underscore that using laboratory or engineering tests to justify fuel efficiency gains for carbon credit calculations has the potential to be extremely misleading. Similarly, most households believe that the stoves reduce cooking time (online Appendix Figure 6). However, we find that, if anything, the stoves increased the time spent cooking evening meals by about three minutes (column 3), although this is not statistically significant at conventional levels.

Finally, we examine the total repairs to stoves. Control households state that they repair their stoves about five times per year. Treatment households made, on average, about 2.4 more repairs to their stove in the last year (column 4), translating to about a half hour of time over the last year (column 5). These two effects are economically large, implying increases of 45 percent and 15 percent, respectively, and are statistically significant.

Despite the fact that GV stoves increase household costs and fail to improve health, households generally report that they are satisfied with the stoves (Table 8). We collected data on the satisfaction with the stoves in the endline survey. On a scale from one to ten, with one being the best, those who obtained an improved stove rate their satisfaction with it at 2.87, with 89 percent of households happy to recommend

[27] The discrepancy between self-reports and actual outcomes has been observed in other contexts as well, and probably reflect social desirability bias, as households do not want to be impolite to people they perceive to be associated with the program (for a discussion, see Kremer et al. 2011). In the stove context, Boy et al. (2000) report that local women in Guatemala stated that the improved cooking stove (plancha) uses less wood than open fire stoves and that this was one of the features that they liked most about the stoves, even though standard measures of the stoves fuel efficiency tests suggested that cooking on a plancha was no more efficient than cooking with an open flame, and may have even required more time to cook.

Table 8—Self-Reported Satisfaction with GV Stoves

	Mean (1)	Observations (2)
Satisfaction with improved stove	2.87	1,200
Would recommend improved stove	0.89	1,192
	Any reasons	
Panel A. Reasons why would recommend		
Requires less time	0.34	1,085
Requires less fuel	0.58	1,085
Food tastes better	0.06	1,085
Less smoke	0.63	1,085
Like the two-pot functionality	0.49	1,085
Easier to clean	0.09	1,085
Better for health of self	0.05	1,085
Better for health of children	0.01	1,085
Pot does not turn black	0.01	1,085
	Any reasons	
Panel B. Reasons why would not recommend		
Requires more time	0.43	140
Requires more fuel	0.43	140
Food tastes worse	0.03	140
More difficult to clean	0.08	140
Have to repair it	0.05	140
More smoke than traditional	0.04	140

Notes: This table provides sample statistics on self-reported satisfaction with the GV improved cooking stoves for those who own one. The satisfaction variable is out of 10, with 1 being the highest level of satisfaction.

the stoves to others. The top reasons for recommending the stoves include that they emit less smoke in the household, the household belief that they require less fuel, the two-pot functionality, and the household belief that they require less time to cook food. The gap between the satisfaction results, which do not commit the households to anything and are likely tainted by social desirability bias, and the consequential health and CO results underscores the limitations of self-reports generally and especially in trying to learn about how individuals value new technologies.

V. Discussion

A. *Relationship with the Literature*

The paper's basic findings differ from the naïve ordinary least squares (OLS) estimates of the impact of improved stoves on health, as well as the conventional wisdom about their benefits in the policy world. Following households for up to four years after they received a subsidized improved stove, we find that the stoves reduce the CO concentration in breath for primary cooks in the first year, but this disappears by the following year. We do not observe any effect on health, neither self-reported nor measured, or on proxies for greenhouse gas emissions.

It is noteworthy that despite the fact that we studied a relatively inexpensive stove model, this study's measured effects are consistent with the findings from the RESPIRE study in Guatemala, which is considered the flagship randomized experiment of a stove program.[28] Online Appendix Table 9 assembles the results from six different RESPIRE papers that summarize the results for outcomes that are similar to those in this paper. The RESPIRE stove resulted in reduced smoke exposure: as panel A demonstrates, for the subsample of women who were tested for emissions exposure, personal $PM_{2.5}$ was about 60 percent lower for the treatment group, relative to the control group, (McCracken et al. 2007), and CO also declined by roughly 60 percent for mothers in the treatment group (Smith-Sivertsen et al. 2009).

Despite these reductions in IAP, the evidence on female health is actually weaker than the way they are typically reported would suggest. Online Appendix Table 9, panel B demonstrates that like our findings, the estimates of the treatment effect on lung functioning are close to zero in magnitude and are not statistically significant (Smith-Sivertsen et al. 2009). There appears to be a modest and statistically significant reduction of blood pressure (panel C), but this was found in the presence of substantial selective attrition (54 percent response rate in the treatment group and 71 percent in the control) and only after adjusting for control variables that increase the magnitude of the point estimate (McCracken et al. 2007). The birth weight results are similar in this respect (panel D). Finally, the evidence on self-reported symptoms is also mixed (panel F). While women in the treatment group experienced a reduction in respiratory symptoms (cough, chronic chough, phlegm cough, phlegm, wheezing, and tightness of chest) the decline was statistically significant only for wheezing. The probability of self-reported sore eyes and headaches was reduced (Diaz et al. 2007).[29]

The effect of the stoves on the incidence of pneumonia and respiratory syncytia virus (RSV) among children are similarly less clear than usually assumed (online Appendix Table 9, panel E). There was no statistically significant difference in MD-diagnosed pneumonia, which is cited as "the primary outcome" in the study. Moreover, out of ten outcomes tracked, only one (severe pneumonia as identified by a field worker) is significant at the 5 percent level in the unadjusted data. It is only after imputing the outcome variable for missing observations that the MD-diagnosed estimates become marginally significant (at the 9 percent level). Even in this adjusted data, the 95 percent confidence interval for the odds ratio (0.59–1.06) *excludes* the figure of 50 percent reduction in pneumonia that is cited by the GACC.

As noted, the RESPIRE experiment had a much larger decline in smoke inhalation than our study, because it was conducted under conditions that approach laboratory tests. For example, the treatment households were visited weekly to ensure that they were using the stove and to provide free repairs. Given the high cost of conducting these kinds of visits, they do not reflect the way that improved stoves are

[28] The stove that we evaluated has some very appealing features in terms of choice of stoves to evaluate in that it is relatively inexpensive to construct, had promising laboratory results, and this family of stoves is currently used by more than 166 million households (World Bank 2011). The RESPIRE stove also has some appealing aspects, although it costs about US$100–150 (Diaz et al. 2007) and the full cost of the very modest health improvements would also have to include weekly inspections and free maintenance.

[29] The effect on headaches was not present in either the six-month or the 12-month follow-ups; it was only present in the 18-month follow-up (Diaz et al. 2007).

typically used, and thus the health results in RESPIRE are likely an upper bound on health benefits of the RESPIRE stove.

There are several potential explanations for these disappointing results. For example, the response function between health and IAP may be highly nonlinear, potentially including thresholds, such that reducing indoor pollution, even by more than 60 percent, at such stunningly high levels may have limited effects. (This is a hypothesis put forward by the *Lancet* review commission on IAP; see Gordon et al. 2014). If this is the case, then it is possible that there is a new stove technology that would reduce indoor pollution exposure sufficiently to lead to health gains, but it is necessary to know what this inflection point should be (the *Lancet* commission recommends achieving the WHO standards, which is very likely impossible without a dramatic shift in the technology and type of fuel used). A related possibility is that the health effects require sustained exposure to air pollution throughout the day, as is the case with the extensive literature that measures the impact of ambient air quality on human health (Chen et. al 2013). In this respect, a problem with the chimney stove is that the smoke remains in the village. In Indian villages, the concentration average levels of ambient $PM_{2.5}$ can exceed 100 $\mu g/m^3$, most of which arises from exfiltration of household smoke (London street air has an average value of 30 $\mu g/m^3$) (Balakrishnan et al. 2013). Alternatively, the overall health status of individuals in very poor countries may be so low that a reduction in indoor pollution would have to be accompanied by other changes to achieve health improvements. In this case, it is possible that stove interventions could be coupled with other health interventions, but again, this would need to be tested.

In addition to RESPIRE, there is emerging evidence from several other randomized evaluations of stoves. Some of the evidence is positive: Bensch and Peters (2012) evaluate an improved cooking stove for about 227 households in Senegal, and find that one year after the stoves are distributed, they are used for about 71 percent of meals in the treatment group, and there are improvements in self-reported fuel and health outcomes.[30] However, the bulk of the evidence is similar to ours. In Mexico, Romieu et al. (2009) conducted a RCT of an improved wood-burning stove, the Patsari stove.[31] Among 668 households, 338 received a free stove. Adherence was low: only 50 percent of treatment women used the stove. There is no significant difference between treatment and control group in terms of respiratory symptoms and lung functions. Evaluating a mud stove in Ghana, Burwen and Levine (2011) show positive effects on self-reported health eight weeks after the distribution, although there is no effect on smoke exposure. However, a year later, only about half of the stoves still appeared to be in use. Similarly, Beltramo and Levine (2010) examine solar stoves in Senegal, and find that six months after the distribution there were no

[30] In Bensch and Peters (2012), there are follow-up visits with households on the stove at the one, two, and seven month marks. Interestingly, in that study, the improved stove moved households from cooking inside to outside.

[31] They describe the stove as follows "The Patsari stove is a new efficient multi-pot wood-burning stove, which has been widely disseminated in rural Mexico. The Patsari stove has shown, in actual field conditions, average reductions of 70 percent in indoor air pollution concentrations, of 56 percent in household fuel consumption, and 74 percent in greenhouse gas emissions compared with open fires."

differences in the amount of time spent cooking near a fire (and only a 1 percent decline in fuel use) since households continued to use the traditional stoves.[32]

B. *Interpretation*

Our paper, as well as the existing literature as a whole, implies that the evidence on improved cooking stoves is significantly less positive than the conventional wisdom would suggest. Further to date, there is not any clear evidence that a large-scale distribution of stoves is likely to be a cost-effective solution to the problem of indoor air pollution. Is this the case because the stoves are just not a good enough technology (even when properly used), or is it because their effectiveness is mediated by improper usage based on households' judgment about the benefits of the technology? Of course, our experiment does not allow us to definitively distinguish between the two possibilities. Nevertheless, our interpretation of the evidence from this study and the literature is that households' beliefs about the potential health benefits play an important and too often ignored role in explaining the existing evidence.

Ultimately, the success of clean stoves depends on household's assessments about whether they are beneficial on net. Our findings are consistent with the possibility that even if the ex ante claims about health benefits were correct, many households came to the judgment that the benefits of proper and frequent stove usage and maintenance were smaller than their costs. On the benefits side of the equation, it is important to note that, when we started this project (and to some extent even now, at least in the public discourse) the consensus was that the clean stoves would have large health impacts; this is what was conveyed to the households at the onset of the projects and throughout. This was done in good faith, since the evidence on reduction in exposure to particulate matter, combined with other evidence on the impact of particulate matter on health, suggested potentially large health effects. The evidence suggests that households "heard" this message, given the answers they gave to the self-satisfaction question.

However, a literature that has emerged since we began this project suggests that the target population may have a relatively low willingness to pay (WTP) for chronic health improvements. Kremer et al. (2011) find low WTP for good water quality in a similarly impoverished population in Kenya. Further, Miller, and Mobarak (2014) suggest that household members (even women) do not value the health benefits associated with reductions in IAP, and that the WTP for clean stoves is very low even when the health impacts are emphasized.

On the cost side, households naturally compared the costs of clean and traditional stoves. It seems reasonable to infer that on average households decided that the clean stoves were more costly than traditional stoves due to greater difficulty in use, higher maintenance costs, space needs, the two-pot feature, etc. So, it is seems at least plausible that most households rationally chose not to regularly use and vigorously maintain the new stoves.

[32] The *Lancet* review cites four more registered ongoing trials (but it excluded all of the existing RCTs except RESPIRE and PATSARI).

Chassang, Padró i Miquel, and Snowberg (2012) provide exactly the right theoretical framework with which to think about this problem. In their framework, the effectiveness of a new technology, tested in an RCT, depends on some unobserved action by the recipient (modeled as an Agent in a principal-agent problem where the experimenter is the principal). If a recipient does not believe that the technology is worthwhile, they will not exercise effort, and an RCT will find that, in reduced form, measured returns will be low. However, it will not be able to distinguish whether this is because "true returns are low or because most agents believe they are low and therefore expend no effort using the technology." They propose the use of "selective trials" to recover the effectiveness of a technology as a function of the agent's willingness to pay: basically, the agent is allowed to exert their preferences for the technology by selecting in or out of the control group at a price. In this setting, this would have required, for example, an elicitation of people's willingness to acquire a stove at different price points, and then the random selection of one price point. We would then have people of different willingness levels pay for the stove in the treatment and control groups, allowing us to observe whether people with higher willingness to pay maintain and use the stove better and have bigger health impacts. This was of course not feasible in the current setting, but it provides an interesting direction for future work.

VI. Conclusion

This study shows that relatively inexpensive stoves, used under real-world conditions, had limited long-run impacts. The stoves reduced smoke exposure for the primary cook in the household in the first year of the study, but after normal use they subsequently had no discernible effect on exposure. The declining effect appears to be the result of stove breakages combined with insufficient investments in maintenance, reductions in the number of meals cooked with clean stoves in good condition, and inappropriate cleaning and use. Further, we failed to find an effect on a range of health outcomes, even in the early years, when indoor air pollution exposure fell.

Our conclusion from this study, as well as other related work, is that there are significant barriers to alternative clean stove models achieving widespread market penetration and usage. The willingness to pay for the health improvement stemming from a reduction in IAP appears to be low. Thus, for a clean stove to achieve widespread market success, it must be cheap to produce and transport to remote areas and also have a very low marginal cost to use, clean, and maintain. Carbon credits could in principle help with the low willingness to pay, but only if stoves actually reduce greenhouse gas emissions in practice. Our study failed to find evidence of any fuel efficiency gain given field conditions, although it is possible that other models may do better in practice.

More broadly, this study illustrates that it is critical to allow for household behavior when evaluating health and environmental technologies to understand their actual effects. Laboratory and laboratory-style field studies are important for understanding the best-case scenario for a technology. However, all technologies must ultimately be used by individuals who reveal their valuations through their usage and maintenance decisions. At a conceptual level, this suggests that the type of "selective trials" recommended by Chassang, Padró i Miquel, and Snowberg (2012) may be particularly suited to evaluation of the possible impact of technologies whose effectiveness is mediated by household behavior. From a policy standpoint, the lesson from this experiment suggests that to understand the full impact of a new technology, households must be followed in realistic conditions (without the support of a research team) for a long period of time (to allow maintenance problems to occur). Otherwise, very real problems like indoor air pollution will remain unsolved.

REFERENCES

Alonso, P. L., S. W. Lindsay, J. R. M. Armstrong, M. Conteh, A. G. Hill, P. H. David, G. Fegan, et al. 1991. "The effect of insecticide-treated bed nets on mortality of Gambian children." *Lancet* 337 (8756): 1499–1502.

Arceo-Gomez, Eva O., Rema Hanna, and Paulina Oliva. 2012. "Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City." National Bureau of Economic Research (NBER) Working Paper 18349.

Balakrishnan, Kalpana, Sambandam Sankar, Jyoti Parikh, Ramaswamy Padmavathi, Kailasam Srividya, Vidhya Venugopal, Swarna Prasad, et al. 2002. "Daily average exposures to respirable particulate matter from combustion of biomass fuels in rural households of southern India." *Environmental Health Perspectives* 110 (11): 1069–75.

Balakrishnan, K., Santu Ghosh, Bhaswati Ganguli, Sankar Sambandam, Nigel Bruce, Douglas F. Barnes, and Kirk R. Smith. 2013. "State and National Household Concentrations of PM2.5 from Solid Cookfuel Use: Results from Measurements and Modeling in India for Estimation of the Global Burden of Disease." *Environmental Health* 12 (1): 77.

Beers, Mark H., and Robert Berkow, eds. 1999. *The Merck Manual of Diagnosis and Therapy*. 17th ed. Rahway, NJ: Merck.

Beltramo, Theresa, and David I. Levine. 2010. "The Effect of Solar Ovens on Fuel Use, Emissions, and Health: Results from a Randomized Controlled Trial." University of California Berkeley, Center for Effective Global Action Working Paper 3.

Bensch, Gunther, and Jörg Peters. 2012. "A Recipe for Success? Randomized Free Distribution of Improved Cooking Stoves in Senegal." Ruhr Economic Paper 325.

Binka, F. N., F. Indome, and T. Smith. 1998. "Impact of Spatial Distribution of Permethrin-Impregnated Bed Nets on Child Mortality in Rural Northern Ghana." *American Journal of Tropical Medicine and Hygiene* 59 (1): 80–85.

Block, Ben. 2013. "India Announces Improved Cook Stove Program." Worldwatch Institute. http://www.worldwatch.org/node/6328 (accessed November 9, 2015).

Burwen, Jason, and David I. Levine. 2012. "A Rapid Assessment Randomized-Controlled Trial of Improved Cookstoves in Rural Ghana." *Energy of Sustainable Development* 16 (3): 328–38.

Boy, Erick, Nigel Bruce, Kirk R. Smith, and Ruben Hernandez. 2000. "Fuel Efficiency of an Improved Wood Burning Stove in Rural Guatemala: Implications for Health, Environment, and Development." *Energy for Sustainable Development* 4 (2): 23–31.

Bruce, Nigel, Lynnette Neufeld, Erick Boy, and Chris West. 1998. "Indoor biofuel air pollution and respiratory health: The role of confounding factors among women in highland Guatemala." *International Journal of Epidemiology* 27 (3): 454–58.

Bruce, Nigel, Rogelio Perez-Padilla, and Rachel Albalak. 2000. "Indoor air pollution in developing countries: A major environmental and public health challenge." *Bulletin of the World Health Organization* 78 (9): 1078–92.

Chassang, Sylvain, Gerard Padró i Miquel, and Erik Snowberg. 2012. "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments." *American Economic Review* 102 (4): 1279–1309.

Chay, Kenneth, and Michael Greenstone. 2003a. "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession." *Quarterly Journal of Economics* 118 (3): 1121–67.

Chay, Kenneth Y., and Michael Greenstone. 2003b. "Air Quality, Infant Mortality, and the Clean Air Act of 1970." National Bureau of Economic Research (NBER) Working Paper 10053.

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): 12936–41.

Currie, Janet, and Matthew Neidell. 2005. "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?" *Quarterly Journal of Economics* 120 (3): 1003–30.

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

Duflo, Esther, Michael Greenstone, Ramond Guiteras, and Thomas Clasen. 2014. "The Short- and Medium-Run Impacts of Clean Water on Diarrhea and Malaria in Rural India." Unpublished.

Duggan, Mark. 2005. "Do New Prescription Drugs Pay for Themselves? The Case of Second-Generation Antipsychotics." *Journal of Health Economics* 24 (1): 1–31.

Dupas, Pascaline. 2011. "Health Behavior in Developing Countries." *Annual Review of Economics* 3: 425–49.

Dupas, Pascaline. 2014. "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment." *Econometrica* 82 (1): 197–228.

Dutta, Karabi, Kyra Naumamoff Shields, Rufus Edwards, and Kirk R. Smith. 2007. "Impact of improved biomass cookstoves on indoor air quality near Pune, India." *Energy for Sustainable Development* 11 (2): 19–32.

Ezzati, Majid, Alan D. Lopez, Anthony Rodgers, Stephen Vander Hoorn, and Christopher J. L. Murray. 2002. "Selected Major Risk Factors and Global and Regional Burden of Disease." *Lancet* 360 (9343): 1347–60.

Global Alliance for Clean Cookstoves. 2010. *Goals*. United Nations Foundation. Washington, DC.

Gordon, Stephen B., Nigel G. Bruce, Jonathan Grigg, Patricia L. Hibberd, Om P. Kurmi, Kin-bong Hubert Lam, Kevin Mortimer, et al. 2014. "Respiratory Risks from Household Air Pollution in Low and Middle Income Countries." *Lancet Respiratory Medicine* 2 (10): 823–60.

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: Dataset." *American Economic Journal: Economic Policy*. http://dx.doi.org/10.1257/pol.20140008.

Jayachandran, Seema. 2009. "Air Quality and Early-Life Mortality: Evidence from Indonesia's Wildfires." *Journal of Human Resources* 44 (4): 916–54.

Johnson, Priscilla, Kalpana Balakrishnan, Padmavathi Ramaswamy, Santu Ghosh, Muthukumar Sadhasivam, Omprakash Abirami, Bernard W. C. Sathiasekaran. 2011. "Prevalence of Chronic Obstructive Pulmonary Disease in Rural Women of Tamil Nadu: Implications for Refining Disease Burden Assessments Attributable to Household Biomass Combustion." *Global Health Action* 4: 7226.

Kandlikar, Milind, Conor C. O. Reynolds, and Andy P. Grieshop. 2009. *A Perspective Paper on Black Carbon Mitigation as a Response to Climate Change*. Copenhagen Consensus Center. Frederiksberg, August.

Kremer, Michael, Jessica Leino, Edward Miguel, and Alix Peterson Zwane. 2011. "Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions." *Quarterly Journal of Economics* 126 (1): 145–205.

McCracken, John P., and Kirk R. Smith. 1998. "Emissions and Efficiency of Improved Woodburning Cookstoves in Highland Guatemala." *Environment International* 24 (7): 739–47.

McCracken, John P., Kirk R. Smith, Anaite Diaz, Murray A. Mittleman, and Joel Schwartz. 2007. "Chimney Stove Intervention to Reduce Long-term Wood Smoke Exposure Lowers Blood Pressure among Guatemalan Women." *Environmental Health Perspectives* 115 (7): 996–1001.

Miller, Grant, and Mushfiq Mobarak. 2014. "Gender Differences in Preferences, Intra-Household Externalities, and the Low Demand for Improved Cookstoves." National Bureau of Economic Research (NBER) Working Paper 18964.

Mobarak, Ahmend Mushfiq, Puneet Dwivedi, Robert Bailis, Lynn Hildemann, and Grant Miller. 2012. "Low Demand for Nontraditional Cookstove Technologies." *Proceedings of the National Academy of Sciences* 109 (27): 10815–20.

Mueller, Valerie, Alexander Pfaff, John Peabody, Yaping Liu, and Kirk R. Smith. 2011. "Demonstrating Bias and Improved Inference for Stoves' Health Benefits." *International Journal of Epidemiology* 40 (6): 1643–51.

Nevill, C. G., E. S. Some, V. O. Mung'ala, W. Muterni, L. New, K. Marsh, C. Lengeler, et al. 1996. "Insecticide-treated bednets reduce mortality and severe morbidity from malaria among children on the Kenyan coast." *Tropical Medicine and International Health* 1 (2): 139–46.

Phillips-Howard, Penelope A., Bernard L. Nahlen, Margarette S. Kolczak, Allen W. Hightower, Feiko O. Ter Kuile, Jane A. Alaii, John E. Gimnig, et al. 2003. "Efficacy of Permethrin-Treated Bed Nets in the Prevention of Mortality in Young Children in an Area of High Perennial Malaria Transmission in Western Kenya." *American Journal of Tropical Medicine and Hygiene* 68 (4): 23–29.

Pitt, Mark M., Mark Rosenzweig, and Nazmul Hassan. 2010. "Short- and Long-Term Health Effects of Burning Biomass in the Home in Low-Income Countries." http://www.brown.edu/research/projects/pitt/sites/brown.edu.research.projects.pitt/files/uploads/iapchild7.pdf.

Registrar General and Census Commissioner, India. 2001. *Census of India*. http://www.censusindia.net/ (accessed November 9, 2015).

Romieu, Isabelle, Horacio Riojas-Rodríguez, Adriana Teresa Marrón-Mares, Astrid Schilmann, Rogelio Perez-Padilla, and Omar Masera. 2009. "Improved Biomass Stove Intervention in Rural Mexico: Impact on the Respiratory Health of Women." *American Journal of Respiratory Critical Care Medicine* 180 (7): 649–56.

Smith, Kirk R. 2000. "National Burden of Disease in India from Indoor Air Pollution." *Proceedings of the National Academy of Sciences* 97 (24): 13286–93.

Smith, Kirk R., John P. McCracken, Lisa Thompson, Rufus Edwards, Kyra N. Shields, Eduardo Canuz, and Nigel Bruce. 2010. "Personal Child and Mother Carbon Monoxide Exposures and Kitchen Levels: Methods and Results from a Randomized Trial of Woodfired Chimney Cookstoves in Guatemala (RESPIRE)." *Journal of Exposure Science and Environmental Epidemiology* 20 (5): 406–16.

Smith-Sivertsen, Tone, Esperanza Diaz, Dan Pope, Rolv T. Lie, Anaite Diaz, John P. McCracken, Per Bakke, et al. 2009. "Effect of Reducing Indoor Air Pollution on Women's Respiratory Symptoms and Lung Function: RESPIRE Guatemala Randomized Trial." *American Journal of Epidemiology* 170 (2): 211–20.

The Global Initiative for Chronic Lung Disease. 2010. "GOLD Spirometry Guide." http://www.goldcopd.org/uploads/users/files/GOLD_Spirometry_2010.pdf (accessed November 9, 2015).

United States Centers for Disease Control and Prevention. 2000. "Growth Charts." http://www.cdc.gov/growthcharts/clinical_charts.htm (accessed November 9, 2015).

World Bank. 2011. *Household Cookstoves, Environment, Health, and Climate Change: A New Look at an Old Problem*. Washington, DC: International Bank for Reconstruction and Development.