Title
Essays in Development, Behavioral, and Health Economics

Author
Harrell, Stephen James

Publication Date
2019

Peer reviewed|Thesis/dissertation

Essays in Development, Behavioral, and Health Economics

by

Stephen James Harrell

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

and the Designated Emphasis

in

Development Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Aprajit Mahajan, Chair

Professor Marco Gonzalez-Navarro

Professor David I. Levine

Fall 2019

Essays in Development, Behavioral, and Health Economics

Abstract

Essays in Development, Behavioral, and Health Economics

by

Stephen James Harrell

Doctor of Philosophy in Agricultural and Resource Economics

with a Designated Emphasis in Development Engineering

University of California, Berkeley

Professor Aprajit Mahajan, Chair

This dissertation combines three empirical studies aimed at addressing one of the developing world's leading causes of death – exposure to household air pollution from using biomass as a cooking fuel – which is responsible for millions of premature deaths annually and contributes to global climate change. The first study tests the effects of a soft commitment device on eliminating biomass use among households in India who have access to a clean fuel, LPG. The second and third studies explore reducing biomass use through a fuel-efficient biomass cookstove among a population in Uganda who does not have access to clean fuels. The second study tests the effects of selling a fuel-efficient cookstove (at the market price) on biomass use. The third study focuses on the methodology behind measuring stove use by comparing four methods of measurement.

The first chapter explores how a commitment mechanism may help households who use a mix of biomass and clean cooking fuels to fully transition to clean fuels. Most Indian households now own an LPG stove and one LPG cylinder. However, many households continue to regularly use indoor biomass-fueled mud stoves (i.e., chulhas) alongside LPG. Focusing on this population in rural Maharashtra, India, this study tests the effects of conditioning a sales offer for a spare LPG cylinder on a soft commitment device requiring initially disabling indoor chulhas. We find that almost all relevant households (>98%) were willing to accept the commitment device. Indoor chulha use decreased by 90% when the sales offer included the commitment device, compared to a 23% decrease without it. If the effects are persistent, this intervention may be one of the most cost-effective means to save lives among tens of millions of Indian households. Using WHO-CHOICE criteria and conservative assumptions, this intervention generates benefits roughly 20 times larger than the costs.

The second chapter, co-authored with Theresa Beltramo, Garrick Blalock, David I. Levine, and Andrew M. Simons, is among the first studies to examine the effects of a fuel-efficient biomass cookstove while selling the stove at market prices. After introducing a fuel-efficient cookstove, fuelwood use and household air particulates declined by 12% and by smaller percentages after adjusting for observer-induced bias, or the Hawthorne effect. These reductions were less than laboratory predictions and fell well short of World Health Organization pollution targets. Even when introducing a second stove, most households continued to use their traditional stoves for most cooking. While any reduction in fuel use and particulate matter was likely beneficial, fuel-efficient biomass cookstoves such as the one used in this study will not be adequate to reach safe levels of household air pollution. Thus, policies that assist consumers to shift to safe fuels such as gas or

electricity—particularly when coupled with policies to disable smoky indoor stoves—should take on increased importance.

The third chapter, co-authored with Theresa Beltramo, Garrick Blalock, Juliet Kyayesimira, David I. Levine, and Andrew M. Simons, compares four methods of measuring stove use: time spent cooking (measured by heat sensors on stoves), number of people cooked for (self-reported), fuel weight used (measured by a scale), and particulate matter concentrations (measured by a monitor). We find statistically significant positive correlations between five out of six of these pairs of measures. While the correlations are positive, the explanatory power of each measure for another is weak. The weak correlations emphasize the importance of using multiple measures to track changes in stove use for both researchers and carbon auditors.

To those I love and to accessing the love within; to the journey of introspection, reflection, and inquiry; and to attempting to alleviate some suffering in the world.

Acknowledgements

In addition to the direct professional support of my mentors and colleagues, this dissertation is surely a product of my upbringing, education, experiences, and interactions with others.

Undoubtedly, my primary advisor, mentor, and friend, David Levine, has played the greatest role in the completion of my PhD. Not only has he provided endless mentoring, but he treats me, and others, as equals. He is focused on improving the lives of the least fortunate and is able to set aside any sense of ego in that pursuit. I have trouble both imagining a better mentor and expressing the depth of gratitude I have for him.

I also thank my advisors Aprajit Mahajan and Marco Gonzalez-Navarro for their thoughtful feedback and kindness throughout the dissertation process. In addition, I thank the many other faculty members who have helped me along the way and have become my friends.

For the first chapter of this dissertation, I especially thank Kirk Smith and Ajay Pillarisetti for the vital roles they played and for all their efforts in addressing related global issues. I thank KEMHRC's research team for carrying out the study and contributing to the research. I thank the Indian Council of Medical Research for funding the study and NSF INFEWS program for additional funding.

For the second and third chapters, I thank CIRCODU for carrying out these studies and my co-authors for their efforts. I thank USAID and Impact Carbon for funding these studies. Additionally, I thank the study participants from all three studies for their participation.

Above all else, I am most grateful for the care, love, and connections I have found with others, beginning with my family and closest friends. This is what I value most and hope to cultivate in all that I do, including this dissertation.

Contents

Chapter 1

Incentivizing elimination of biomass cooking fuels through a commitment mechanism and a spare LPG cylinder

I. Introduction

In India, approximately 480,000 deaths and 16 million lost DALYs (disability-adjusted life years) occur per year due to exposure to household air pollution (HAP) from continuing use of biomass as a cooking fuel (GBD 2017). The Indian government is addressing this problem through distribution of liquefied petroleum gas (LPG) stoves and initial fuel cylinders at a reduced price to poorer households through the Pradhan Mantri Ujjwala Yojana (Ujjwala or PMUY) program.[1] PMUY largely solves the problem of access to clean fuels, yet positive health gains will not be realized unless households nearly completely abandon biomass use (Smith and Pillarisetti 2017; Johnson and Chiang 2015).[2] Unfortunately, there is a significant portion of households that use both LPG and indoor biomass-fueled mud stoves (i.e., chulhas) on a regular basis (Gould and Urpelainen 2018; Anenberg et al. 2013). Jain et al. (2018, Fig. 42) find that about 50% of rural households across six states use both LPG and biomass for cooking. If that share generalizes, then roughly 90 million rural households in India use both stoves regularly.[3] Our study focuses on this population of mixed users by only including rural households who own an LPG stove, have purchased at least three LPG refills in the past nine months, and continue to use an indoor chulha.

Among this population, the primary suggested reasons for using both LPG and biomass for cooking include (Jain et al. 2018): (1) supply issues – there is a gap in time between when a household's LPG cylinder becomes empty and when they receive a refill, (2) information problems – respondents were unaware of the harms of biomass use, and (3) behavioral barriers – households have the habit, routine, and preference for using a chulha.

The supply issue – of lacking constant access to LPG – results in a return to biomass use for cooking while the household waits for a refill. Middle and upper-income households have solved this problem by owning a second, or spare, cylinder. That is, when the first cylinder runs out, they have a second cylinder to use while they wait for the refill of their empty cylinder. To acquire a second cylinder, households must pay approximately $20 (INR 1,450) as a security deposit.[4] Although the recurring cost of LPG fuel may be affordable to our population of interest, this fixed cost may be a significant barrier in acquiring a second cylinder.

[1] See: http://www.pmujjwalayojana.com/faq.html.
[2] This is due to the nonlinear exposure-response relationship between HAP and health outcomes (Johnson and Chiang 2015).
[3] The 2011 India Socio Economic and Caste Census states there are 180 million rural households. See: https://timesofindia.indiatimes.com/india/Census-2011-data-released-10-key-highlights/articleshow/47923276.cms.
[4] See: https://www.bankbazaar.com/gas-connection/indane-gas-new-connection-price-and-charges.html.

Despite the increase in LPG use in India, there is low awareness of the negative health effects of burning biomass for cooking (Jain et al. 2018). Understanding the negative impact burning biomass has on health may lead to a decrease in biomass use and an increase in alternative fuel use (i.e., LPG use).

Our basic intervention addresses supply and information problems through a sales offer of a free trial (followed by installment payments) of a second, or spare, cylinder coupled with health educational messages on the harms of using biomass for cooking with chulhas. The combination of a free trial and installment payments is used because a previous study (Levine et al. 2018) found that it increased fuel-efficient biomass cookstove sales by a factor of ten. In our study, we also experimentally identify, through a second study arm, the effects of one strategy to address behavioral barriers – conditioning the sales offer on disabling the indoor chulha.

Despite expressed dissatisfaction from households with using chulhas, as long as the indoor chulha is still present and functional, it is unlikely that households will fully transition away from using it. To explore this, we test the effects of conditioning the sales offer on a soft commitment device – requiring the initial disabling of the indoor chulha by either dismantling it or filling it with mud or pebbles. This is a soft commitment device because the consequences are largely psychological – the process can be undone within an hour if the chulha was filled with pebbles, or within a half day if it was dismantled, and there is no other associated penalty with subsequent use (i.e., households still received the sales offer if they rebuilt the chulha during the study as long as they initially disabled it). Given the ease of undoing the commitment, the effectiveness of the strategy is uncertain.

Behavior change issues are well-known in the public health and technology adoption literature, especially in development economics (e.g., Kremer, Rao, and Schilbach 2019). Commitment devices have been used as one strategy to encourage behavior change (e.g., Kremer, Rao, and Schilbach 2019). In this literature, past research has focused on the effects of both hard commitment devices – those that have economic penalties or rewards, and soft commitment devices – those that have largely psychological consequences (Bryan, Karlan, and Nelson 2010). Hard commitment devices have found large effects among a variety of outcomes including worker performance (Kaur, Kremer, and Mullainathan 2015), savings (Kast, Meier, and Pomeranz 2018), and exercise (Royer, Stehr, and Sydnor 2015). Soft commitment devices have also found significant effects in areas including savings (Thaler and Benartzi 2004) and educational progress (Himmler, Jackle, and Weinschenk 2017). One study, Karlan and Linden (2018), compared a hard and soft commitment savings account and found that the soft commitment savings account caused stronger increases in savings for educational supplies.

For our context, more closely related literature involves government policies using soft commitment devices that remove or inhibit use of certain goods. These studies have found positive, but somewhat mixed results. Mikhed, Scholnick, and Byun (2017) find that the removal of slot machines from neighborhood bars in Alberta, Canada reduced personal bankruptcies filed by close neighbors. Bernheim, Meer, and Novarro (2016) find that consumers in relevant US states increase their liquor consumption in response to extended Sunday on-premises (i.e., at restaurants and bars) sales hours, but not in response to extended off-premises sales hours (i.e., through liquor stores).

A key distinction of this paper from past research on government policies is that participants voluntarily remove the good (i.e., the indoor chulha) instead of the government changing barriers to use of certain goods. Because the commitment here is voluntary, there may be a larger effect for those that agree to the terms. On the other hand, given the ability to easily re-enable the indoor chulha, households may choose to use it shortly after initially disabling it.

Specific to our context, Pillarisetti et al. (2019) explored related issues in a non-randomized sample of a specific population – pregnant women. That study loaned 200 households in rural Maharashtra, India a second LPG cylinder for the duration of the study in order to ensure they have constant access to LPG. They asked, but did not require, households to disable their indoor chulha during the study. Surprisingly, 65% disabled it. At the end of the study, households were asked to either purchase the second cylinder or return it: 85% chose to purchase it. These previous findings provide suggestive evidence that a soft commitment device to disable the indoor chulha in exchange for a second LPG cylinder loan may lead to significant effects on decreasing indoor chulha use, and uptake of a second LPG cylinder. This provides motivation for our study, which aims to eliminate (or at least greatly decrease) use of biomass fuels among households that own one LPG cylinder and use LPG and an indoor chulha regularly.

Our study finds that almost all (>98%) of the relevant households in our study site were interested in the sales offer regardless of whether it was conditional on disabling the indoor chulha. When the sales offer was conditional on initially disabling the indoor chulha, chulha use decreased by 90% (based on minutes used as determined by data from temperature-logging stove use monitors, or SUMs). Without this requirement, chulha use decreased by only 23%. We find no statistically significant change for either treatment group in LPG use (based similarly on minutes used). For both treatment groups, 80% purchased the second cylinder at the end of the study.

These results imply a high willingness of households to agree to a soft commitment device (i.e., disabling the indoor chulha) that leads to a large decrease of 77% in indoor chulha use. If the effects are persistent, this intervention may be one of the most cost-effective means to save lives among tens of millions of Indian households. Using WHO-CHOICE criteria and conservative assumptions, this intervention generates benefits roughly 20 times larger than the costs. If the Indian government wishes to reduce deaths from household air pollution substantially, the intervention we studied may provide a highly cost-effective model.

The remainder of the paper is organized as follows. Section II describes the experimental setting and study design. Section III describes the specification used. Section IV reports the results. Lastly, Section V concludes and discusses implications for policy makers and researchers.

II. Experimental setting and study design

A. Study site and participant selection

Our study was conducted in a subset of the Junnar block of Pune district among a population of approximately 6,000 individuals living in 1,200 households. The study was conducted by KEMHRC Pune, which has worked extensively in the area for decades conducting health-related research. This location was chosen due to KEMHRC's experience with the study area and the high percentage of households that use biomass for cooking. According to LPG distributors in the area, approximately

70% of households owned only one LPG cylinder at the time of the study and hence were potentially eligible for participation.

Working with Accredited Social Health Activists (ASHAs), we identified all households in a subset of the study area (371 in total) that own exactly one LPG cylinder, have purchased at least three refills in the past nine months, and own an indoor chulha. We then approached a random subset of 189 of the eligible households, explained the study design (while verifying that they satisfy the eligibility criteria), and invited them to participate in the study. Of the 189 households, three did not agree to participate (before treatment status was known) because they did not want to be assigned to the study arm that is required to disable the indoor chulha in order to receive the sales offer. After obtaining written informed consent, each household was randomly assigned to one of three study arms.

B. Study design

All households that agreed to participate in the study (N=186) were randomly assigned to one of the following three intervention arms:

1. Treatment 1 (N=62): receive a six-week free trial of a second LPG cylinder (with the right to return it during the free trial) followed by four installment payments collected over ten additional weeks. This offer also included a detailed set of health educational messages to explain the harms of using biomass for cooking;
2. Treatment 2 (N=62): same offer as Treatment 1 but with a requirement to disable or move their indoor chulha in order to receive the sales offer;
3. Control group (N=62): no special sales offer or health messages but provides access to purchase a second cylinder, as is commonly available in the area.

The study included a four-week baseline period in addition to the sixteen-week endline period. A baseline survey was conducted at the beginning of the study to gather household demographics. Households were visited every two to four weeks to monitor stove usage and to deliver health messages.[5] Due to some of the originally recruited households not satisfying the eligibility criteria of owning an indoor chulha, approximately half of the households began the study in December 2018 and the other half began in March 2019. The study concluded in July 2019.

Two sets of health messages were used in the study. The first is a flipchart (see Appendix 1) that focuses on the harms to health of using biomass with related images. The second is a video[6] that focuses on different scenarios in which households may be inclined to use biomass (e.g., during festivals, when they run out of LPG, or taste preferences) and then suggests potential solutions.

C. Outcomes

As the primary goal of our intervention was to eliminate use of biomass for cooking, our primary outcome is indoor chulha use. LPG use is also measured as a secondary outcome. For chulha and

[5] Note, the control group did not receive health messages.
[6] To view the video, see: http://kemhrcvadu.org/index.php/projects/ongoing-projects/14-sample-data-articles/189-impact-of-clean-cooking-fuel-on-health.

LPG stove use, outcomes include time spent using each stove per day, cooking events per day on each stove, and a binary indicator for whether or not each stove was used each day. We also measure uptake rates of the sales offer for each treatment group and rates of purchasing the second cylinder at the end of the study.

D. Measurements

Thermocouple-based stove use monitors (SUMs) manufactured by Wellzion (model number SSN-61, Xiamen, Fujian, China) were used on both chulhas and LPG stoves to measure stove usage in all households in the study. Although we do not account for the effects of the presence of SUMs on behavior (e.g., as Thomas et al. 2016 does), we assume these effects are equal across each study group and thus do not bias the comparison of study groups. In treatment group 2 (in which indoor chulhas were disabled), we placed SUMs on all filled in chulhas to monitor if households used them. In cases where households completely destroyed the indoor chulha, we monitored if they rebuilt the chulha at each household visit (without any associated penalty). If they did rebuild it, we placed SUMs on it. For households that completely destroy their indoor chulha, indoor chulha use is assigned zero for all relevant days and then weighted proportional to rates for households in which we have valid SUMs data.[7]

SUMs record temperature readings every 10 minutes and can stay powered for weeks at a time. In order to analyze stove use, we first need to convert SUMs data – in the form of temperature time series data – into metrics of stove use. For this analysis, we used a slightly modified version of the FireFinder algorithm, part of the open-source SUMSarizer R package[8] maintained by Geocene. Firefinder builds upon previous SUMs detection algorithms (Pillarisetti et al. 2014, Piedrahita et al. 2016, Ruiz Mercardo et al. 2012) to determine approximate time spent cooking on each stove (in minutes)[9], number of cooking events per day (count), and days of any stove use (binary).

III. Specification

The following ordinary least squares regression is used to determine the effects of each treatment on traditional biomass and LPG stove use:

$$Y_{it} = \alpha_i + b_1T_{1i} + b_2T_{2i} + b_3post_t + \beta_1\, T_{1i}{*}post_t + \beta_2\, T_{2i}{*}post_t + \varepsilon_{it}$$

where Y_{it} is the stove use outcome (i.e., minutes per day, cooking events per day, or binary indicator for any use per day) for household i at time t (t=0 if pre-treatment, t=1 if post-treatment), α_i are household fixed effects (which are necessary due to the autocorrelative nature of daily measurements from the same household), T_{1i} is a dummy variable equal to one if in treatment group 1, T_{2i} is a dummy equal to one if in treatment group 2, $post_t$ is a dummy variable equal to one if occurring after treatment,

[7] Note, approximately 30% of SUMs data is missing due to over-heating, malfunctioning, and a shortage of SUMs. The missing data is spread evenly across each study group arm.
[8] See: https://github.com/Geocene/sumsarizer.
[9] Note, interpreting time spent cooking between stove types should be done with caution as there is a much slower decay in heat on a chulha compared to that of an LPG stove.

and ε_{it} is the error term. The coefficients of interest are β_1 and β_2 which, due to the exogeneity of treatment status, are the causal effect of being in treatment groups 1 and 2, respectively, post-treatment. As the data may be correlated at levels above our unit of randomization, which is at the household level, standard errors are clustered at the village level.

For the binary outcomes of accepting the sales offer and purchasing a second cylinder we will run a two-sample t test for equality of proportions.

IV. Results

A. Summary Statistics and Randomization Tests

Table 1 shows baseline summary statistics. The average household had 4.6 people and owned one indoor chulha. Most (64%) of households were below the poverty line. The average respondent was 40 years old and had nine years of education. On average, households used the indoor chulha for 53 minutes per day (averaging over all days) and used it on 44% of days. They also averaged 101 minutes of LPG stove use per day and used LPG on 77% of days.

Table 2 shows balance tests for covariates comparing the control group, treatment 1, and treatment 2. The 16 covariates were not jointly statistically significant in predicting treatment arm ($\chi^2 = 31$, P = 0.42). Of the 48 possible comparisons, four were statistically significant. Treatment 1 had almost one year less of education than the control group (8.37 vs 9.28). Treatment 2 had slightly fewer households with electricity access than the others (95% vs 100%), used LPG on 10% more days than treatment 1 (83% vs 73%), and used LPG for about 25 more minutes (i.e., 25% more) per day than either treatment 1 or the control group (119 vs. 94 vs. 89). Fortunately, the core baseline covariates of chulha use (daily, number of cooking events per day, and minutes per day) were balanced across the study groups.

B. Initial and final sales offer uptake

Before being randomly assigned to a study group, households consented to participate in the study. Of 189 households approached, three (less than 2%) did not consent to participate. After randomization, all who consented accepted their initial sales offer. The three who declined consent are included in the analysis as one in each study group. Thus, 98% of treatment 1 and 2 each accepted the relevant initial sales offer (Table 3). Post-randomization, all treatment 2 households complied with their sales offer and initially disabled the indoor chulha. This implies that households did not perceive disabling their indoor chulha as a costly action compared to the benefit of receiving a free trial of a second cylinder. Of these participants, 53% destroyed their indoor chulha and 47% filled it with mud or pebbles. All 53% did not rebuild their chulha throughout the study. No treatment 1 or control group household destroyed or filled their indoor chulha throughout the study. At the end of the study, roughly 80% of treatment 1 and 2 purchased the second cylinder after the free trial while 20% returned it. No participants from the control group purchased a second cylinder during the study.

C. Effects of sales offer on biomass and LPG use

Next, we analyze our primary outcome, the causal impact of each sales offer on biomass use (Table 4). In the pre-intervention period, households on average use the indoor chulha for 53 minutes per day (our most precise measure), for 0.8 cooking events per day, and on 44% of days. After receiving the sales offers, based on daily minutes used, treatment 1 decreases indoor chulha use by 23% (-12 min, 95% CI = -23 to -0.88, p<0.05) while treatment 2 decreases indoor chulha use by 90% (-48 min, 95% = -70 to -26, p<0.01. Difference from treatment 1 P = 0.002).[10] Thus, conditioning the second cylinder sales offer of a free trial + installment payments + health messages with the requirement to initially disable the indoor chulha leads to an additional 77% decrease in minutes of indoor chulha use. In regard to the number of cooking events per day and percent days of any use, we find no change for treatment 1 households while we find a 61% decrease in both measures for treatment 2 households (-0.48 cooking events per day, 95% CI = -0.831 to -0.127, p<0.01; -27% of days, 95% CI = -40% to -14%, p<0.01).

Lastly, we analyze the causal impact of each sales offer on LPG use (Table 5). In the pre-intervention period, households on average use the LPG stove for 100 minutes per day (our most precise measure), for 2.6 cooking events per day, and on 77% of days. After receiving the sales offers, we find no statistically significant change for either treatment group in minutes or event per day of using LPG. We do however find a small and weakly statistically significant decrease in percent days using LPG among Treatment 2 by 6.5% (p<0.10).

These results of LPG use are puzzling as one would expect to see an increase in LPG use to compensate for the decrease found in chulha use. Potential reasons for this lack of reduction include households may have increased their efficiency of using LPG by using both burners on the LPG stove (and perhaps because they were no longer simultaneously using their indoor chulha), households may have shifted indoor chulha use to outdoor chulha use (which we did not observe), or sampling error.

V. Discussion and Conclusion

Household air pollution from using biomass as a cooking fuel is a significant contributor to ill-health in India, approaching five hundred thousand deaths per year (GBD 2017). The PMUY program helped enable most households to access a safe cooking fuel (LPG). However, many households regularly use indoor biomass along with LPG.

Our study finds a potential solution for many of these households – a sales offer for a spare LPG cylinder coupled with the requirement to initially disable the indoor chulha. We find high demand for this sales offer (98%). We also find a huge effect, a 90% reduction, in indoor biomass use. Additionally, we find that the soft commitment device is vital to this very large reduction; without this commitment device, the sales offer results in just a 23% reduction in indoor biomass use. These results imply a high willingness of households to agree to a soft commitment (i.e., disabling the indoor chulha) that leads

[10] Note, during endline we find a puzzling overall decrease in any daily chulha use and no. of daily uses, but there is no change in the more precise measure of daily minutes used (see table 4 row 1).

to a large decrease of 77% in indoor chulha use. Future research is needed to determine how well the effects last outside of a four-month period and what their impacts are on household air pollution.

If the effects are persistent, this intervention may be one of the most cost-effective means to save lives among tens of millions of Indian households. Jain et al. (2018, Fig. 42) find that about half of rural households across six states use both LPG and biomass for cooking. If that share generalizes, then 90 million rural households in India use both stoves regularly.[11] While some homes rely on smoky stoves for heat or other purposes, it is likely our intervention could reduce household air pollution in the vast majority of these 90 million homes.

We next discuss cost-effectiveness. Nearly half a million deaths per year and 16 million lost DALYs are attributable to household air pollution in India. Our relatively low-cost intervention may decrease this burden. One method of evaluating the cost-effectiveness of health interventions is using WHO-CHOICE criteria. Interventions that avert one DALY for less than average per capita income or region are considered very cost-effective. Using this approach, Tripathi and Sagar (2019) find that the total lost economic value from household air pollution is roughly INR 69,000 (~USD $1,000) per household that cooks with biomass. That economic value is presumably lower for households that also cook on LPG, but the majority of health harms are persistent whenever biomass cooking remains common (Smith and Pillarisetti 2017; Johnson and Chiang 2015). The cost of this intervention is on the order of less than USD $10 per household to cover overhead, transportation of LPG cylinders, hiring ASHAs to deliver health messages, and ensuring households initially disable the indoor chulha. Thus, even with very conservative estimates, the benefits of this intervention far exceed the costs.

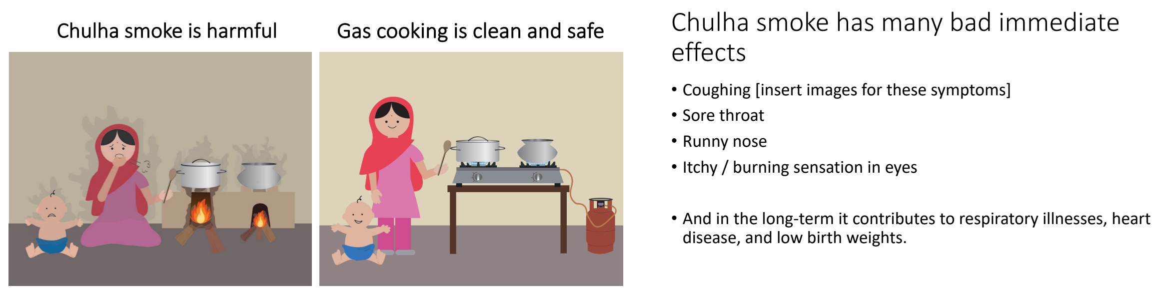
As an illustration of a conservative estimate on benefits vs. costs, suppose the indoor chulha has only 20% of the economic value of removing a biomass stove in other settings (because the households already have LPG, and some will rebuild their indoor chulha). This would result in a benefit of $200 per household. In regard to costs, suppose this sales offer costs four times that of a normal delivery, or $1.60. Suppose ASHAs receive a similar amount as they do for promoting other health related products, or $1 per home visit.[12] Assume uptake of the sales offer is two-thirds of what we observed (65%, not 98%), then the ASHA cost per household that accepts the offer is $1.50. Assuming total overhead costs of the intervention are double the field costs (or $6.20 per household), then the total cost per household is $9.30. With these conservative assumptions, this intervention generates benefits roughly 20 times larger than the costs (i.e., $200 in benefits vs. $10 in costs per household). Even if the cost-benefit ratio is off by a factor of 10, this intervention is still very cost-effective.

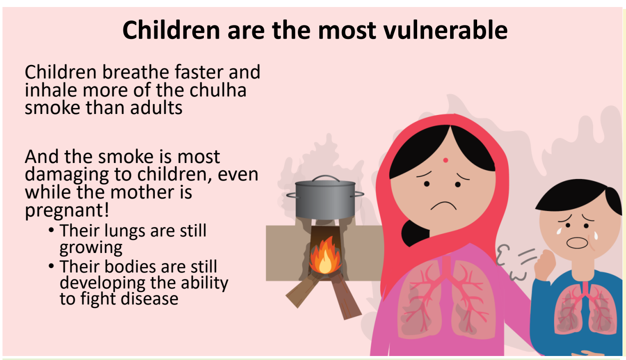
If the Indian government wishes to substantially reduce deaths from household air pollution, an intervention that targets disabling indoor chulhas is likely to be important. The intervention we studied may provide a highly cost-effective model.

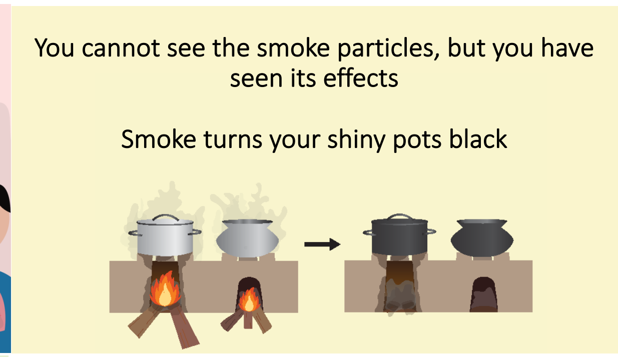
[11] The 2011 India Socio Economic and Caste Census states there are 180 million rural households. See: https://timesofindia.indiatimes.com/india/Census-2011-data-released-10-key-highlights/articleshow/47923276.cms.
[12]ASHAs receive roughly $1 for promoting household toilets. See Table 1 in the following: https://www.intrahealth.org/sites/ihweb/files/files/media/performance-based-payment-system-for-ashas-in-india-what-does-international-experience-tell-us-technical-report/PerformanceBasedPaymentSystemASHAsIndiaReport.pdf

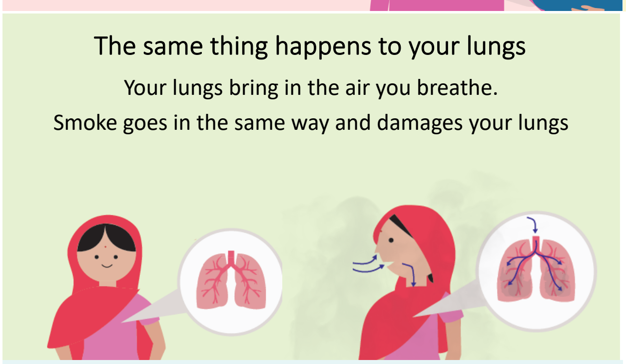
Appendix

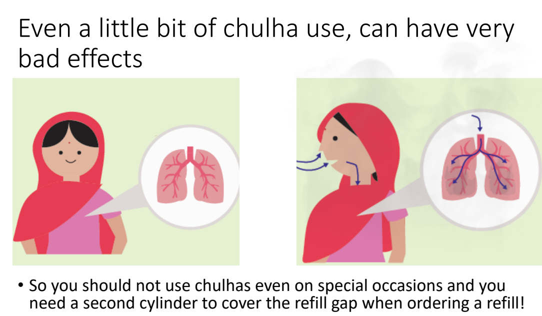
Appendix 1 – Health message 1

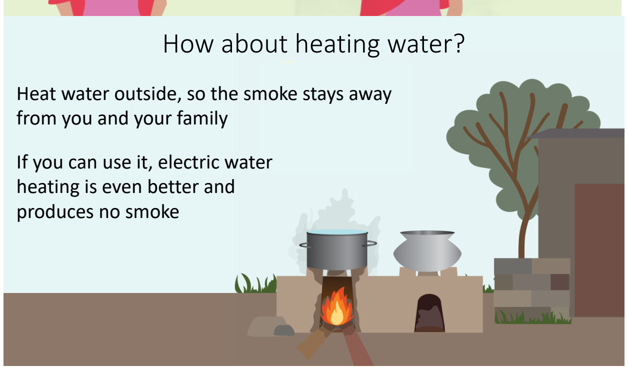












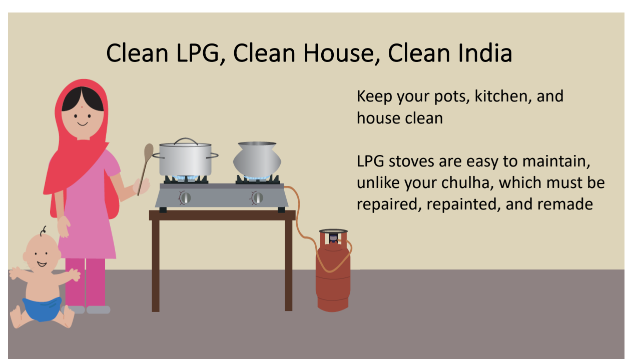


Table 1: Baseline summary statistics

Statistic	Mean	St. Dev.	Min	Max	N
Household size	4.63	1.73	2.00	12.00	185
Respondent is primary cook (share)	0.86	0.35	0.00	1.00	185
Respondent age	39.68	12.45	19.00	72.00	185
Female respondent (share)	1.00	0.00	1.00	1.00	185
Years of education	8.93	2.99	0.00	15.00	185
Below poverty line (share)	0.64	0.48	0.00	1.00	185
Electricity access (share)	0.98	0.13	0.00	1.00	185
Number of mud stoves owned	1.04	0.19	1.00	2.00	185
Received LPG through PMUY (share)	0.06	0.24	0.00	1.00	185
Years owned LPG	7.01	5.43	0.17	24.00	185
Days used mud stove (share)	0.44	0.36	0.00	1.00	171
Days used LPG stove (share)	0.77	0.29	0.00	1.00	182
No. of daily mud stove uses	0.79	0.78	0.00	4.23	171
No. of daily LPG stove uses	2.55	1.36	0.00	5.88	182
Mud stove daily minutes used	52.69	61.05	0.00	276.21	171
LPG stove daily minutes used	101.12	72.44	0.00	519.38	182

Note: To be included in the study, households were required to own at least one mud stove and have an LPG connection with one LPG cylinder. This data includes baseline stove usage measured four weeks pre-intervention. Days used mud/LPG stove (share) refers to the share of days using the respective stove during the four-week baseline period. PMUY refers to a government program in India that distributes LPG connections to below poverty line families.

Table 2: Baseline summary statistics and balance of covariates (treatment groups combined)

	Covariate	C mean	C SD	T mean	T SD	Difference	p-value	N
1	Household size	4.46	1.41	4.71	1.87	-0.25	0.31	185
2	Respondent is primary cook (share)	0.85	0.36	0.86	0.35	-0.01	0.85	185
3	Respondent age	39.08	11.74	39.98	12.83	-0.89	0.64	185
4	Female respondent (share)	1	0	1	0	0		185
5	Years of education	9.28	2.45	8.76	3.21	0.52	0.22	185
6	Below poverty line (share)	0.72	0.45	0.60	0.49	0.12	0.11	185
7	Electricity access (share)	1	0	0.98	0.15	0.02	0.08	185
8	Number of mud stoves owned	1.02	0.13	1.05	0.22	-0.03	0.21	185
9	Received LPG through PMUY (share)	0.03	0.18	0.07	0.26	-0.04	0.23	185
10	Years owned LPG	6.86	5.78	7.09	5.27	-0.22	0.80	185
11	Days used mud stove (share)	0.48	0.36	0.42	0.34	0.05	0.35	171
12	Days used LPG stove (share)	0.77	0.28	0.78	0.29	-0.01	0.90	182
13	No. of daily mud stove uses	0.84	0.81	0.77	0.73	0.07	0.58	171
14	No. of daily LPG stove uses	2.40	1.21	2.59	1.37	-0.19	0.35	182
15	Mud stove daily minutes used	51.05	52.87	53.15	60.93	-2.10	0.82	171
16	LPG stove daily minutes used	88.99	50.25	104.93	75.19	-15.94	0.09	182
	χ^2 test for joint significance						0.15	

Note: C and T refer to control and treatment groups respectively. There are 61 control households and 124 treatment households. PMUY refers to a government program in India that distributes LPG connections to below poverty line families.

Table 3: Initial and final offer uptake

	Offer uptake		**Purchased 2^{nd} cylinder**		**Sample size**
	%	p-value	%	p-value	N
Control	NA		0%		63
Treatment 1	98%		82%		63
Treatment 2	98%		81%		63
Control vs Treatment 1		NA		0.000***	
Control vs Treatment 2		NA		0.000***	
Treatment 1 vs Treatment 2		1.000		1.000	

Note: *p<0.1; **p<0.05; ***p<0.01
Households consented to participate in the study before being randomly assigned to a group. All who agreed to participate in the study accepted the sales offer they were assigned. They were offered 500 INR (USD $7.30) for participation. Three households did not consent to participate, so we count those declining households as one in each group. This accounts for the 98% in initial offer uptake. The control group received no special sales offer. Treatment 1 received a free trial of a second LPG cylinder. Treatment 2 received a free trial contingent on disabling the indoor mud stove: 53% destroyed it and 47% filled it with mud or pebbles. All 53% did not rebuild their mud stove throughout the study.

Table 4: Regressions of daily indoor mud stove use and treatment group

	All columns include household fixed effects		
	Daily minutes used	No. of daily uses	Any daily use
	(1)	(2)	(3)
Endline	-6.07	-0.26^{***}	-0.15^{***}
	(3.88)	(0.07)	(0.03)
Treatment1*Endline	-11.92^{**}	-0.02	-0.003
	(5.64)	(0.12)	(0.04)
Treatment2*Endline	-47.57^{***}	-0.48^{***}	-0.27^{***}
	(11.21)	(0.18)	(0.07)
Baseline mean	52.69	0.79	0.44
Observations	16,343	16,342	16,343
No. of household FEs	166	166	166
R^2	0.33	0.37	0.44

Note: *p<0.1; **p<0.05; ***p<0.01
Treatment 1 received a free trial of a second LPG cylinder. Treatment 2 received the free trial contingent on disabling the indoor mud stove. The study includes approx. four weeks of pre-endline data and 16 weeks of endline data. There is some missing data due to monitors over-heating or malfunctioning. Col (3)'s outcome variable refers to an indicator if any use occurred that day with the stove. Standard errors are clustered at the village level.

Table 5: Regressions of daily LPG stove use and treatment group

	All columns include household fixed effects		
	Daily minutes used	No. of daily uses	Any daily use
	(1)	(2)	(3)
Endline	1.51	−0.08	-0.04^{**}
	(6.98)	(0.15)	(0.02)
Treatment1*Endline	−5.33	−0.10	0.01
	(6.09)	(0.17)	(0.02)
Treatment2*Endline	−12.61	−0.22	-0.05^{*}
	(11.12)	(0.23)	(0.03)
Baseline mean	101.12	2.55	0.77
Observations	18,032	18,031	18,032
No. of household FEs	179	179	179
R^2	0.36	0.36	0.33

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Treatment 1 received a free trial of a second LPG cylinder. Treatment 2 received the free trial contingent on disabling the indoor mud stove. The study includes approx. four weeks of pre-endline data and 16 weeks of endline data. There is some missing data due to monitors over-heating or malfunctioning. Col (3)'s outcome variable refers to an indicator if any use occurred that day with the stove. Standard errors are clustered at the village level.

Chapter 2

The Effects of Fuel-Efficient Cookstoves on Fuel Use, Particulate Matter, and Cooking Practices: Results from a Randomized Trial in Rural Uganda

I. Introduction

Almost 3 billion people cook with wood, charcoal, and dung using traditional cookstoves (Bonjour et al. 2013). These stoves cause environmental degradation (Bailis et al. 2015), global climate change (Ramanathan and Carmichael 2008), and an estimated four million deaths per year (Lim et al. 2012). Truly safe cooking likely requires clean fuels such as gas or electricity. Unfortunately, most people who cook with solid fuel lack an affordable and consistent supply of gas or electricity (Lewis and Pattanayak 2012; Rehfuess et al. 2010). In the short to medium term, fuel-efficient cookstoves that use less solid fuel than traditional stoves may reduce these environmental and health problems.

We experimentally examined the effects of a fuel-efficient cookstove, the Envirofit G3300 wood-burning stove, on wood use, household air pollution, and cooking behaviors in rural Uganda. Our work builds on important antecedents and extends previous literature in three key ways: (1) households purchased the new stove at the market price; (2) we provided households with a second fuel-efficient stove to see if a second cooking surface would limit stove-stacking; and (3) we adjusted for observer-induced bias, or the Hawthorne effect.

The first studies to document the relationship of stove usage, household air pollution, and human health were conducted in Kenya (Ezzati and Kammen 2001, 2002; Ezzati, Saleh, and Kammen 2000) and Guatemala (Smith et al. 2006; Smith et al. 2011; Smith-Sivertsen et al. 2009). More recently, Hanna, Duflo, and Greenstone (2016) examined the link between stove usage and household air pollution in India and found reductions in smoke inhalation in the first year, but no changes over longer periods. They suggested that the fade-out was due to a lack of stove maintenance by users. Bensch and Peters (2015) examined a stove designed to reduce fuelwood consumption in rural Senegal and found reductions in fuelwood use, smoke emissions, and smoke-related disease symptoms. Pillarisetti et al. (2014) examined stove usage in a sample of pregnant women in India and found that users experimented with the fuel-efficient stove at first, but that the use of the new stove declined over time. Moreover, by one year after introduction, the sampled households used traditional stoves for 75% of their cooking.

Similar to the studies of Hanna, Duflo, and Greenstone (2016) and Bensch and Peters (2015), we measured stove use in the short term (a year or less) and over the long term (a 3.5 year follow-up). These two previous studies measured health outcomes (documented by medical personnel or self-reported). In contrast, we measured household level particulate matter (PM2.5) concentrations.

Particulate matter concentrations have been directly linked to health problems in numerous studies (Chay and Greenstone 2003; Currie and Walker 2009; Smith-Sivertsen et al. 2009). Due to their small size (2.5 µg or less), these particles can reach deep into the lungs and are the best single indicator of risk for many respiratory-related diseases (Chowdhury et al. 2007).[1] Similar to Pillarisetti et al. (2014), we used unobtrusive temperature sensors to measure detailed household stove use over time.[2] However, unlike Pillarisetti et al. (2014), we introduced random variation in the assignment of when the stoves were delivered to causally examine the effects of the introduction of a fuel-efficient stove.

Our study extends previous literature in three important ways. First, we examine cooking behaviors among households that were willing to purchase the new stove at market prices (and perhaps, therefore, value the stove more highly).[3] Because our results come from users who paid the market price for the fuel-efficient stove, our sample mimics those that would be most likely to purchase such a stove. There is a long-standing debate whether developing countries should charge for health improving products (latrines, mosquito bed nets, deworming medications, chlorine tablets, etc.) or if they should be distributed for free (Ashraf, Berry, and Shapiro 2010; Cohen and Dupas 2010; Dupas 2014; Fischer et al. 2019). A key part of this debate is the question of how usage of the product varies depending on the price paid. Generally, cookstoves have been given for free or highly subsidized in previous cookstove usage studies. Our study adds a new data point to quantify usage for users who paid market price for their cookstoves.

A second innovation in our study was that, after measuring stove usage when households had one fuel-efficient stove, we provided all households with a second fuel-efficient stove. Common cooking practice in the study area involved cooking with two pots simultaneously (e.g., rice and beans, or steaming bananas and cooking gravy). Stove stacking (the simultaneous use of the fuel-efficient stove and the traditional cooking technology) has been mentioned as a challenge to completely switching to fuel-efficient stoves (Masera, Saatkamp, and Kammen 2000; Pillarisetti et al. 2014; Ruiz-Mercado et al. 2011). This non-experimental intervention allowed us to examine how important the lack of a second cooking surface was for continued use of the traditional stove.

A third innovation of our study was that we adjusted for observer-induced bias, or the Hawthorne effect. The Hawthorne effect has been mentioned as a potential source of bias in numerous cookstove studies (Bensch and Peters 2015; Ezzati, Saleh, and Kammen 2000; Pillarisetti et al. 2014; Smith-Sivertsen et al. 2009). By collecting sensor data both when observers were and were not present, we were able to measure and remove the source of this observer-induced bias.[4]

[1] According to Pope III et al. (2002), each 10 µg/m³ increase in long-term exposure to fine particulate matter is associated with approximately a 4%, 6%, and 8% increase in the risk of all-cause cardiopulmonary and lung cancer mortality, respectively.

[2] These stove usage monitors were pioneered by Ruiz-Mercado, Canuz, and Smith (2012).

[3] Among these similar studies, Hanna, Duflo, and Greenstone (2016) distributed highly subsidized stoves (users paid US$0.75 for a US$12.50 stove), while Bensch and Peters (2015) and Pillarisetti et al. (2014) distributed stoves for free. Studies primarily focusing on the public health benefits of cookstoves typically distribute the cookstoves for free. For example, the randomized exposure study of pollution indoors and respiratory effects (RESPIRE) in Guatemala (Smith et al. 2006; Smith-Sivertsen et al. 2009), the Cooking and Pneumonia Study in Malawi (Mortimer et al. 2017), and the research on emissions, air quality, climate, and cooking technologies in Northern Ghana (REACCTING) study (Dickinson et al. 2015).

[4] This adjustment removes the bias from when observers were present compared to when no observers were present. We acknowledge that it is possible that the sensors themselves could have induced different behavior, however we feel that given the small size of the sensors (about the size of a coin) and the length of tracking (about six months) that the sensor was not salient enough to make a big difference in sustaining atypical cooking behaviors.

During the weeks when wood use and particulate matter were measured, we found that the randomized early introduction of the first fuel-efficient stove reduced wood use by 11.6% and particulate matter by 12.0%. Once both fuel-efficient stoves were introduced, wood use declined by 26.7% and particulate matter by 10.0%. However, we also found that participants cooked more on the fuel-efficient stoves and less on three-stone fires when observers were present, and that participants reversed these changes once observers left (Simons et al. 2017). When adjusting for this observer-induced bias, we found that the randomized early introduction of the first fuel-efficient stove may have only reduced wood use by 1.7% and particulate matter by 0.3%. Once both fuel-efficient stoves were introduced, after adjusting for the Hawthorne effect, we found wood usage may have declined by 2.5% compared to the baseline; however, particulate matter may have increased[5] (an increase of 18.3% compared to the baseline).

Households used the new stoves more hours per day than the usage of the three-stone fires declined. The increase in total hours of stove usage blunted reductions in fuel use and household air pollution. At the same time, cooking on multiple surfaces most likely increased the utility of the cooks. It appears that cooks used each stove for the foods that fit it best. For example, low-heat simmering of rice, beans and unripe bananas was done on three-stone fires, and making sauces and boiling water for tea was done on the fuel-efficient stove. In the longer term (3.5 years), we found lower rates of disrepair than Hanna, Duflo, and Greenstone (2016).[6] Nevertheless, as in their study, we found low longer-term usage of the fuel-efficient stove.

Concerning related environmental problems, our findings suggest fuel-efficient cookstoves similar to the one used in our study and setting have, at best, marginal effects. The 12% reduction in fuel use (upon introduction of the first fuel-efficient stove) may generate small reductions in deforestation and carbon dioxide emissions, at least in the short term (though these reductions dissipated over the length of our study).

Concerning related health problems, the 12% reduction in particulate matter left the air 14 times more polluted than the World Health Organization (WHO) standard of 25 $\mu g/m^3$ (World Health Organization 2006). Thus, if clean air is a high priority, our findings suggest it is important to help consumers shift to safe fuels such as gas or electricity and to find ways to encourage them to disable or move their smoky stoves outdoors.

[5] Note that the introduction of the second fuel-efficient stove was not experimentally identified, and the difference in changes in particulate matter and wood use could have been due to a variety of factors, such as weather changes (i.e., wet wood burns less efficiently).

[6] This pattern makes sense as Hanna, Duflo, and Greenstone (2016) examined local artisan-built mud stoves, while the stoves used in our study were commercially manufactured from metal. The manufacturer (Envirofit Inc.) stated its stoves would last up to ten years. See: https://www.envirofit.org/.

II. Experimental Setting and Data

A. Background and Site and Stove Selection

We selected the Mbarara region of Uganda because it is rural, almost all families cooked on a traditional three-stone fire, households spent significant time gathering firewood or purchased firewood, and the local government was supportive of our work. In pre-experimental discussion groups, we confirmed that there was no active fuel-efficient cookstove intervention in the region, and that families spent significant time gathering wood (approximately 10–20 hours per week).

Most participants farm matooke (starchy cooking banana), potatoes, and millet and raise livestock. Prior to our experiment, almost all families cooked on a traditional three-stone fire (97%), usually located within a separate cooking hut. Most (62%) households had totally enclosed kitchens with no windows, while 38% had semi-enclosed kitchens with at least one window. Almost all cooking occurred in the detached cooking hut.

We implemented a series of companion studies in rural areas of the Mbarara District in southwestern Uganda from February to September 2012, focusing on the adoption of fuel-efficient stoves. These studies analyzed the household purchase decision, and they found that relieving liquidity constraints by allowing additional time for payments (Beltramo et al. 2015b) and providing a free trial with time payments allowed users to learn about the stoves' fuel savings properties (Levine et al. 2018) and greatly increased purchase rates (for example, from 5% to 57% in our setting in rural Uganda). We also examined how social networks affected purchasing (Beltramo et al. 2015a).

We marketed the Envirofit G3300 wood-burning stove, made by Envirofit International Inc. (Ft. Collins, CO, USA) (see Figure 1 for images of a traditional three-stone fire and the Envirofit G3300). This stove achieves relatively efficient fuel combustion by channeling airflow into the fire and directing heat upward through an insulated cylinder to the cooking surface. These design innovations allow fuel to burn at a controlled rate and enable more complete combustion than a three-stone fire. Emissions testing of the Envirofit G3300 in a controlled laboratory setting found average reductions in carbon monoxide (CO) of 65%, particulate matter reductions of 51%, and a reduction in fuel wood use of 50% compared with a three-stone fire (see Figure 2 for a copy of the emissions and performance report).

Before selecting the Envirofit G3300, we conducted a feasibility study that tested four different models of fuel-efficient stoves among households within the study zone.[7] The feasibility study included three focus groups and one town hall style meeting, which included a total of 85 participants. This study found that the participants preferred the Envirofit G3300. Additionally, during the feasibility study it was apparent that most households used two cooking points on most days. This finding informed our experimental design to distribute a second Envirofit to each household to give cooks the ability to completely substitute away from the use of traditional three stone fires.

[7] The full feasibility study report can be found here: https://www.cleancookingalliance.org/binary-data/CMP_CATALOG/file/000/000/153-1.pdf.

B. Selection of Study Participants

In the first stage of the experiment, we randomly selected 12 parishes (units of government administration covering about 4,000–6,000 people), to receive a traditional full upfront payment sales offer and 14 parishes to receive a sales offer of a one-week free trial followed by four equal weekly time payments (see Levine et al. 2018). Within each parish, we recruited a local point person with the help of local government officials. We asked each focal point person to gather roughly 60 people together for a public sales meeting on a specified day. We did not tell the point person which sales offer his or her parish would receive.

At the sales meeting, participants completed a questionnaire that focused on household cooking and basic socioeconomic indicators. After this, the study team presented the Envirofit G3300, discussed the stove's features such as fuel savings and reduced pollution relative to traditional three-stone fires, gave a cooking demonstration, and presented the terms of the randomly selected sales offer. While the Envirofit was not commercially available in this region prior to our experiment, we sold it for the same retail price (40,000 Ugandan shillings [~US$16]) that it was selling for in parts of the country where it was available. We used the randomized assignment of the sales offer by parish as the identifying assumption, as used by Levine et al. (2018), to examine the barriers to purchase. In the current paper, to examine how often people used their stoves, our identification strategy was based on randomly assigning the timing of when purchasers received their Envirofit (we call them early buyers and late buyers). In each of the 14 parishes with the sales offer of a free trial plus time payment, we randomly selected 12 of the purchasing households for stove usage tracking. Therefore, all participants who had their stove usage tracked received the same sales offer at the extensive margin, and all participants fully paid for the stove according to the terms of the sales offer (one-week free trial, followed by four equal payments totaling 40,000 shillings).

Households were eligible to have their stove usage tracked if they mainly used wood as a fuel source, regularly cooked for eight or fewer persons (so that their cooking pots could fit on the Envirofit), someone was generally home every day, and cooking was largely done in an enclosed kitchen. In each parish, more than 12 households met these criteria and agreed to join the study; therefore, among those that agreed, we randomly selected 12 households per parish to track with the stove use monitors (SUMs). We then randomly assigned each of these 12 households to be an early buyer or late buyer. We asked both early and late buyers if they would agree to have SUMs immediately placed on their traditional three-stone fires (all agreed). We used the randomly assigned time of Envirofit delivery (early buyers vs. late buyers) as the identifying assumption for the causal claims made in this paper.

After participants consented to participate in the usage study, all existing three-stone fires were affixed with SUMs. Then, approximately four weeks after the SUM data collection began, the early buyers' group received their first Envirofit stove. Approximately four weeks after that, the late buyers received their first Envirofit stove.

Based on earlier studies (e.g., Pillarisetti et al. 2014; Ruiz-Mercado et al. 2011, 2013) and our feasibility-study, we anticipated that many households would use both their three-stone fire and their Envirofit. One motivation for this is that common cooking practices in the area require two simultaneous cooking pots (for example, for rice and beans, or for matooke and a sauce), and the Envirofit heats only one pot. We were interested in whether having a second fuel-efficient stove would substantially end stove stacking. Thus, approximately four weeks after late buyers received their first Envirofit, we surprised both groups with the gift of a second Envirofit stove.

In short, during the first study wave, both early and late buyers had only three-stone fires; in the second study wave, early buyers had one Envirofit, along with their three-stone fires, but late buyers only had three-stone fires; in the third study wave, both groups of buyers had one Envirofit; and in the fourth wave, both early buyers and late buyers had two Envirofits. See Table 1 for the steps of the experimental rollout. We tracked stove temperatures for approximately 18 weeks (May–September 2012). Each household had as many as two three-stone fires and two Envirofit stoves monitored with SUMs. By the end of the study, numerous SUMs had been lost or burned up; therefore, after we delivered the second Envirofit stove, we encountered a shortage of SUMs, so we focused measurement on both Envirofits and the primary three-stone fire.

C. SUMs

We installed small, inexpensive, and unobtrusive SUMs to record stove temperatures.[8] Ruiz-Mercado et al. (2008) initially suggested using SUMs to log stove temperatures, and various studies have used that method (Mukhopadhyay et al. 2012; Pillarisetti et al. 2014; Ruiz-Mercado et al. 2013). We installed SUMs on two Envirofits and two three-stone fires in each household when possible (recall that by the end of the study, numerous SUMs had been lost or burned up; therefore, only a few secondary three-stone fires were measured when all users had two Envirofits).

Throughout the study, field staff recorded about 2,400 visual observations of whether a stove was in use (on/off) when they visited homes. Also, we examined the temperature data immediately before and after the 2,400 visual observations of stove use. After understanding how temperature patterns changed at times of observed stove use, we developed an algorithm to predict cooking behaviors for the wider dataset of 1.7 million temperature readings during which we did not have visual observations. By "cooking," we mean that the algorithm predicts stove use, not necessarily that a cook is standing above the fire and actively working on a meal. Our algorithm would likely detect "cooking" in cases of banking hot coals for the next meal, and while this is not a formal act of cooking, it does burn wood and increase particulate matter in the kitchen. This process, detailed in Simons et al. (2014a), allowed us to unobtrusively and inexpensively track daily stove usage on a large sample of households throughout the study. Appendix provides additional details on placing SUMs, the process of converting temperature readings into measures of predicted cooking, and documents that SUMs attrition was random.

[8] The SUMs used for our project, iButtons™ manufactured by Maxim Integrated Products, Inc., are small stainless steel temperature sensors about the size of a small coin and the thickness of a watch battery. Our SUMs recorded temperatures up to 85°C with an accuracy of +/– 1.3°C. For additional details see: http://berkeleyair.com/services/stove-use-monitoring-system-sums/. The SUMs cost approximately US$16 each. They recorded a temperature data point every 30 minutes for 6 weeks in a household before needing minimal servicing from a technician to download the data and reset the device.

D. Kitchen Performance Tests and Particulate Matter Monitoring

We performed standard kitchen performance tests (KPTs) (Bailis, Smith, and Edwards 2007) in each household to measure the quantity of fuel wood used, record detailed food diaries of what households cooked, and measure household air pollution before any Envirofits were distributed, that is, when early buyers had one Envirofit and when both groups of buyers had two Envirofits. The KPT lasted approximately 72 hours and involved daily visits by a small team of researchers who weighed wood and collected food diaries, which record cooking and stove usage over the previous 24 hours. Households were asked to only use wood from a specific pile so that the team could determine the change in weight over each day. In the food diary, households recorded what foods were cooked for each meal.

During household visits, we also monitored household air pollution. Residential combustion of solid fuels in developing countries is a significant source of pollutants that harms both the climate and health (Bond et al., 2004; Smith et al, 2004). Roughly 10%–38% of the carbon contained in fuels is not completely combusted when used in simple cooking technologies (Zhang et al., 2000). The carbon that is not converted into CO_2 is instead emitted as products of incomplete combustion (PICs) that contain potent health-damaging pollutants. We measured household level particulate matter (PM2.5) concentrations over the same 72 hours of the KPT. To measure PM2.5, we used the University of California, Berkeley (UCB) Particle and Temperature Sensor, which is a small, portable data logging device (a modified commercial smoke detector) that uses an optical scattering sensor to measure real-time PM2.5 concentrations.[9]

E. Long-Term Stove Usage

We revisited households approximately 3.5 years after they initially received their Envirofit stoves. The survey team made quick, unannounced, observation visits in November 2015 to see whether Envirofit stoves were still in use. The purpose of the visits was to observe which stoves were in use at the time of the visit, examine Envirofits and three-stone fire locations for obvious signs of use (smoke stains, black soot, etc.), and ask a series of short qualitative consumer satisfaction questions about the different stove types. We observed 82% (137 of 168) of the households.

III. Specification

We analyzed wood usage (kg/day), daily household air pollution (PM2.5) concentrations, and stove usage. Recall that there were four study waves with different levels of stove ownership: (1) households that had two three-stone fires; (2) early buyers who had received an Envirofit and late buyers who had only their three-stone fires; (3) both groups of buyers that had one Envirofit; (4) both groups of buyers that had received a second Envirofit. Due to budgetary constraints, we could only run KPTs at phases

[9] The UCB Particle Monitor User Manual (Berkeley Air Monitoring Group and Indoor Air Pollution Team, School of Public Health, University of California 2010) details how to use these sensors.

(1), (2), and (4). Thus, for outcomes measured in KPTs (wood usage, PM2.5), the regression specification using data from study waves (1), (2), and (4) was as follows:

$$(1)\quad Y_{ipt} = \alpha_{ip} + b_0 * T_i + b_1 * Early\_have\_Envirofit_t + b_2 * Both\_have\_two\_Envirofits_t + \beta_1 (T_i * Early\_have\_Envirofit_t) + \beta_2 (T_i * Both\_have\_two\_Envirofits_t) + \epsilon_{ipt}$$

where Y_{ipt} is daily wood use or daily PM2.5 concentrations for household i for parish p in study wave t, α_{ip} are fixed effects for each household, $Early\_have\_Envirofit_t$ and $Both\_have\_two\_Envirofits_t$ are dummies for the study wave, and T_i is a dummy equal to one if, in the early treatment group, ϵ_{ipt} is a residual that is clustered by the parish * study wave but is assumed to be independent and identically distributed (i.i.d.) within a parish and study wave. The coefficients of interest are β_1 (the effect of being in the early buyer group during the study wave [2], or the effect of owning an Envirofit while the comparison group has only three-stone fires), and β_2 (the effect of being in the early buyer group during study wave [4], or the effect of owning your first Envirofit for approximately 4 weeks longer than the comparison group when both groups own two Envirofits).

We also ran this equation without household fixed effects, but our preferred specification included them. The household fixed effect controls for unobserved characteristics of the household, such as the talent and cooking style of the household cook, and structural features of the kitchen, such as windows or ventilation. Because particulate matter has extreme positive outliers, we analyzed the natural log of PM2.5 (as is typical in studies that examine PM2.5). We also top and bottom coded PM2.5 at the 2nd and 98th percentiles, and top coded wood usage at the 98th percentile.

For stove usage, we had data for both during and between the three weekly periods when we measured wood usage and PM2.5. Thus, the regression specification for the SUM usage data was:

$$(2)\quad Y_{ipt} = \alpha_{ip} + b_0 * T_i + b_1 * Early\_have\_Envirofit_t + b_2 * Both\_have\_Envirofit_t + b_3 * Both\_have\_two\_Envirofits_t + \beta_1 (T_i * Early\_have\_Envirofit_t) + \beta_2 (T_i * Both\_have\_Envirofit_t) + \beta_3 (T_i * Both\_have\_two\_Envirofits_t) + \epsilon_{ipt}$$

where Y_{ipt} is daily three-stone fire or Envirofit usage derived from SUM readings for household i for parish p in study wave t, α_{ip} are fixed effects for each household, $Early\_have\_Envirofit_t$, $Both\_have\_Envirofit_t$, and $Both\_have\_two\_Envirofits_t$ are dummies for the study wave, and T_i is a dummy equal to one if, in the early treatment group. ϵ_{ip} is a residual that may be clustered by the parish * study wave but is assumed to be i.i.d. within a parish and study wave. The coefficients of interest are β_1 (the effect of being in the early buyer group during study wave [2], or the effect of owning an Envirofit while the comparison group has only three-stone fires), β_2 (the effect of being in the early buyer group during study wave [3], or the effect of owning your first Envirofit for approximately 4 weeks longer than the comparison group which also owns one Envirofit), and β_3 (the effect of being in the early buyer group during study wave [4], or the effect of owning your first Envirofit for approximately 4 weeks longer than the comparison group when both groups own two Envirofits).

A. Accounting for the Hawthorne Effect

Wood usage and PM data are only from field technicians' visits during the approximately 72-hour KPT measurement week. In a companion paper (Simons et al. 2017), we found that there was a significant Hawthorne effect during those weeks.[10] In an attempt to account for this effect, we calculated differences in stove usage between weeks when observers were present and weeks when they were not present and adjusted wood and PM2.5 measures as follows.

Let the subscript group = early or late buyer, and let the superscript wave = the experimental wave (i.e., [1] households with two three-stone fires; [2] early buyers with an Envirofit and late buyers only with three-stone fires; [3] both groups of buyers with one Envirofit; and [4] both groups of buyers with two Envirofits). Our estimate of wood usage adjusted for the Hawthorne effect was:

$$(3)\quad \Delta\text{Adj\_Wood}_{\text{group}}^{\text{wave}} = \Delta\text{TSF\_Hours}_{\text{group}}^{\text{wave}} * \left(\frac{\text{TSF\_Wood}}{\text{hour}}\right) + \Delta\text{ENV\_Hours}_{\text{group}}^{\text{wave}} * \left(\frac{\text{ENV\_Wood}}{\text{hour}}\right)$$

ΔTSF_Hours and ΔENV_Hours are the differences in hours cooked due to the Hawthorne effect on the three-stone fire (Envirofit) among those that own Envirofits. TSF_Wood per hour is wood consumption from the first KPT (when no one had an Envirofit) divided by cooking on the three-stone fires during those days. We did not have any periods when households only had Envirofits. Thus, we used the laboratory results (Figure 2) indicating that ENV_Wood per hour is half that of a three-stone fire.

For PM concentrations, we followed the same technique, and the Hawthorne-adjusted PM2.5 generated for each group of buyers was:

$$(4)\quad \Delta\text{Adj\_PM2.5}_{\text{group}}^{\text{wave}} = \Delta\text{TSF\_Hours}_{\text{group}}^{\text{wave}} * \left(\frac{\text{TSF\_PM2.5\_Generated}}{\text{hour}}\right) + \Delta\text{ENV\_Hours}_{\text{group}}^{\text{wave}} * \left(\frac{\text{ENV\_PM2.5\_Generated}}{\text{hour}}\right)$$

TSF_PM2.5_Generated per hour is calculated by dividing PM2.5 concentrations by three-stone fire use from the first kitchen performance test (when no one had an Envirofit). ENV_PM2.5_Generated per hour is from laboratory results (Figure 2).

Because we had sensor-based usage metrics that covered all weeks of the experiment, the estimates for changes in cooking behaviors (hours cooked per day on three-stone fires and Envirofits) from Eq.

[10] We compared stove usage in KPT weeks when observers were present with stove usage in adjacent weeks with no observers and found that participants increased usage of Envirofits by about 3.0 hours per day and decreased usage of the primary three-stone fires by about 1.8 hours per day during the endline KPT (when households owned two Envirofits), but then reverted to previous usage patterns once the observers left (Simons et al. 2017). Also, see Garland, Gould, and Pennise (2018) for an additional example of observer-induced behavioral differences in stove use during kitchen monitoring periods.

(2) were not likely affected by the observer-induced behavioral response.[11] However, because technicians were in homes to measure wood usage and PM2.5, we adjusted for the Hawthorne effect by using Eqs. (3) and (4).

IV. Results

A. Summary Statistics and Randomization Tests

Table 2 shows baseline summary statistics and balance tests for covariates. Randomization between early buyers and late buyers was successful. Only one difference among the 20 covariates was (weakly) statistically significantly different than zero. Participants who randomly received their Envirofits early had a higher value of assets (US$1,158 vs. US$905) ($p=0.08$). Control households used approximately 9.3 kg of daily wood, had an average PM2.5 reading of 414.3 $\mu g/m^3$ in their kitchens, and cooked for about 6.2 people.

Households used their first Envirofit about 4.3 hours per day and their second Envirofit about 2.9 hours per day (Table 3).

B. Effects of Envirofits on Fuel Use and Pollution

We began by analyzing the causal impact of the introduced Envirofit stove on wood usage (Table 4) during our experiment. In the pre-intervention period, the control group used about 9.3 kg of wood/day (Table 2, column 1); these usage rates fell when the early group had one Envirofit (-1.9 kg/day, $p<0.01$, Table 4, column 1) and when both groups had two Envirofits (-2.5 kg/day, $p<0.01$, Table 4, column 1), but there were no statistically significantly different rates of reduction for those that had received their Envirofit in the early group. In our preferred specification, with household fixed effects (column 2), the early receipt of an Envirofit was causally associated with a change of about -1.1 kg/day, ($p<0.1$). This reduction in wood consumption was a modest reduction of about 12% from the pre-intervention control group wood usage level. When all owned two Envirofits, both groups reduced their wood usage by about 2.5 kg/day ($p<0.01$) or 27%, relative to the pre-intervention control group, with no statistically significant difference between groups.

In Table 5, we present the causal effects of the introduction of Envirofit stoves on household air pollution concentrations. Pre-intervention, the control group had a daily concentration of PM of about 414 $\mu g/m^3$ (Table 2, column 1). In our preferred specification with household fixed effects (Table 5, column 2), the introduction of the first Envirofit reduced PM concentrations by 12% ($p<.01$) compared to the control group. When both groups had two Envirofits, both groups reduced PM by about 10% ($p<0.1$) with no difference between groups. That is, having the first Envirofit longer did not result in detectably different pollution levels once both groups had received two Envirofits.

[11] Observers (technicians) were only present in households in three 72-hour periods over the 18 weeks that sensors measured stove usage.

C. Effects of Envirofits on Cooking Behaviors

Next, we examined the effects of the introduction of Envirofits on daily time spent cooking on the existing three-stone fires. We had stove usage data for much longer periods than the three kitchen measurement periods. We estimated how the daily hours cooked on each stove varied over the entire 18 weeks of the study period (Table 6, based on Eq. 2). Figure 3 summarizes stove usage by study phase. A weekly time series of stove usage is shown Figures 4 and 5.[12]

Total usage on both three-stone fires was 12.7 hours per day by the control group in the sample of all weeks prior to Envirofit introduction. In our preferred specification (Table 6, column 2), the causal estimates were that the introduction of the first Envirofit reduced cooking on three-stone fires by about 3.7 hours per day ($p<0.01$). This was a reduction of about 30% from the control group prior to the introduction of the first Envirofit.

When late buyers received their first Envirofit (Table 6, column 2), we saw a reduction in use of the three-stone fires among late buyers by 3.1 hours per day ($p<0.01$) (about 25%); however, at the same time, we saw an increase in three-stone fire use of about 2.9 hours per day ($p<0.01$) (about 23%) in the early buyers (who had owned their Envirofits about 4 weeks longer than the late buyers). It is unclear why these differed in direction, though one possibility is that, after initial experimentation with the Envirofit, the early group had decided to use their three-stone fires more, while the late group continued to experiment with the new Envirofit. This difference appears to have resolved itself once both groups received their second Envirofit (Table 6, column 2), as combined use of the three-stone fires declined by about 5.2 hours per day ($p<0.01$, with no statistically significant difference if households received their first Envirofit earlier or later). This was a reduction of about 41% in three-stone fire use once both Envirofits were introduced. In short, even with two Envirofit stoves, most households continued to use their three-stone fires regularly.

D. Adjusting for the Hawthorne Effect

To adjust for this effect, we calculated the change in three-stone fire and Envirofit hours cooked in the measurement week compared to all weeks.[13] To do this for three-stone fires, we ran the regression for the effect of the Envirofit on hours cooked on three-stone fires, but restricted the sample to only observations during the measurement week (Table 7). The difference of the coefficients between Table 6 (all weeks) and Table 7 (only measurement weeks) was the delta three-stone fire hours used in Eqs. (3) and (4). To calculate the change in hours cooked on Envirofits, we ran similar regressions, but instead used hours cooked on the Envirofit as the dependent variable (Table 8 [all weeks] and Table 9 [measurement weeks]).

[12] See Appendix Figures A1 and A2 for the daily time series of stove use by early and late buyers, respectively.

[13] Note that this is one option for addressing the Hawthorne effect. As this is not a methodological paper, we only show this option, but we realize that other options are reasonable (e.g., only use one week before/after observers are present to adjust estimated use). Thus, we add the caveat that this method is only a rough estimation of the Hawthorne effect on differences in wood use and particulate matter.

Use of three-stone fires fell by 6.4 hours per day when the first Envirofit was delivered, when only looking at the week when the KPTs were performed (Table 7, column 2), versus 3.7 hours per day over the entire period with sensors (Table 6, column 2). Usage of the Envirofit was roughly 3.8 hours per day when the first Envirofit was delivered, when only looking at the kitchen measurement week (Table 9, column 3), versus 1.5 hours per day over the entire period with sensors (Table 8, column 3). This reduction in three-stone fire use and increase in Envirofit use was anticipated because the measurement weeks had the Hawthorne effect resulting from the daily visits of our enumerators (Simons et al. 2017).

Thus, we adjusted for the 2.6 hours per day increased use of three-stone fires and 2.4 hours per day (Table 10) decreased use of one Envirofit outside of the measurement week using Eqs. (3) and (4). This adjustment yielded a smaller estimated reduction in wood use: 1.7% (Table 11, first panel) as opposed to the unadjusted reduction of 11.6% (Table, 4, column 2). We also found a smaller reduction of PM2.5: 0.3% (Table 11, second panel) instead of the unadjusted reduction of 12.0% (Table 5, column 2).

Next, we calculated the Hawthorne adjustment for the periods when participants had two Envirofits. Use of three-stone fires fell 10.2 hours per day when participants had two Envirofits during the measurement week (Table 7, column 2), versus 5.2 hours per day during the entire period with sensors (Table 6, column 2). Use of the Envirofits was 6.8 hours per day during the measurement week (Table 9, columns 3 and 5), versus 3.7 hours per day during the entire period with sensors (Table 8, columns 3 and 5).[14] Therefore, we adjusted for the 5.1 hours per day increased use of the three-stone fires and 3.1 hours per day (Table 10) decreased use of two Envirofits outside of the measurement week using Eqs. (3) and (4). The estimate of daily wood use changed from an unadjusted reduction of 26.7% to a reduction of 2.5% after the adjustment (Table 11, panel one). The estimate of daily PM2.5 concentrations changed from an unadjusted reduction of 10.0% to an increase of 18.3% after the adjustment (Table 11, panel two).

E. Long-term Usage

We made unannounced visits to measure stove usage approximately 3.5 years after the initial Envirofit stoves were distributed. Approximately 82% of the original households were home when we visited.

At the exact moment our enumerators arrived, about 48% (66 out of 137) of the households were actively cooking (Table 10). Among those, only 9% (6 out of 66) were cooking with an Envirofit stove. Enumerators asked the 131 households that were not cooking on the Envirofit when enumerators arrived if they could inspect their Envirofit to see obvious signs of use, such as black soot or fresh ashes in the stove (Figure 6 shows an example of a stove with obvious signs of use). Among those households, 65% had an Envirofit with obvious signs of use, 17% had Envirofits stored that were clearly not being used, 2% had Envirofits that were still in perfect condition (essentially never used), 8% said their Envirofit was damaged and disposed of, and a final 8% said they had given the stove

[14] We calculated total Envirofit cooking as the sum of cooking on the first Envirofit plus the cooking on the second Envirofit individually, because only about 60% of the households had any combined readings from both SUM devices during the final measurement week.

away. Next, enumerators asked households to see their second Envirofit to determine if it had signs of use. Among this sample, 25% had a second Envirofit with obvious signs of use, 11% had their second Envirofit stored with limited signs of use, 9% had a second Envirofit that had never been used, 38% reported they had given the second Envirofit away as a gift, and 16% said the second Envirofit was damaged and they disposed of it.

Among all households visited (N=137), 23% reported that they still used both Envirofits, 50% said they used only one Envirofit, and 27% said they had stopped using Envirofits completely.

Enumerators also asked all households if they had to purchase a stove today, would they purchase an Envirofit. Among respondents, 79% said they would purchase an Envirofit, and 15% said they would not purchase an Envirofit, with the remaining households unsure. Given that the share that stated a willingness to repurchase was greater than the share using the Envirofit, we suspected this self-report was biased.

Enumerators then asked open-ended response questions as to the reasons for those hypothetical purchase decisions. The most popular responses among those that would buy another Envirofit were that the stove saved fuel and reduced household time collecting fuel, the stove cooked fast, the stove was easily portable, and the stove produced less smoke than a three-stone fire. Among those that said they would not purchase another Envirofit, the most popular responses were that the preparation of firewood was difficult for Envirofits (needed smaller pieces of wood than a three-stone fire), the stove did not simmer food, the stove was too small for the household's cooking needs, it was hard to prepare some traditional meals on the stove, and the stove was hard to light.

F. Rebound Effects

Rebound effects occur when improvements in energy efficiency make consuming energy less expensive and therefore encourage increased consumption of energy (see review in Sorrell, Dimitropoulos, and Sommerville [2009]). While we did not have fuel cost data to formally estimate a rebound effect, we examined stove use graphically, as shown in Figure 3, which suggested the presence of a rebound effect. When households first received an Envirofit, they reduced three-stone fire usage. However, by the end of our tracking period, Envirofit usage had increased more than three-stone fire use had decreased. The aggregate time all stoves were in use increased by about 20% throughout the period that we tracked stove temperatures.

V. Discussion and Conclusion

This study was the first randomized trial that collected detailed stove usage metrics among households that paid market prices for their stoves. We found a slight reduction in wood use (-11.6%) and PM2.5 concentrations (-12.0%) after the introduction of one Envirofit, but this reduction mostly vanished if we adjusted for the Hawthorne effect.

Despite our selection of a sample that paid market price for their fuel-efficient stove, it did not appear that usage rates of the new stove were markedly different than studies that offered highly subsidized stoves. For example, in Pillarisetti et al. (2014), which also used temperature sensors to track detailed stove level usage, households received fuel-efficient stoves for free and ended up using their traditional stoves about 75% of the time and the introduced fuel-efficient stove about 25% of the time. Our results were very similar, with roughly 67% of cooking done on the three-stone fires and 33% on fuel-efficient surfaces by the end of our study. Hanna, Duflo, and Greenstone (2016) did not gather stove use monitor data; however, their conclusion was that fuel-efficient stove use was enough to reduce indoor air pollution in the initial phase of their experiment, but that in the longer term, poor maintenance of the stoves led to an elimination of the air pollution benefits. Our results were similar, except that, in our follow-up, it did not appear that a lack of stove durability was the cause of limited stove use.

A second innovation in our study was to see if households would fully switch from the traditional smoky cookstove, if given a second Envirofit. Despite the second fuel-efficient cooking surface, households continued to mostly use the traditional cookstove. Almost all households used both three-stone fires and fuel-efficient stoves in daily cooking. It appeared that households used the fuel-efficient stove to heat things that cook relatively quickly, such as boiling water to make tea and sauces. They preferred three-stone fires for low-heat cooking, such as simmering dishes like beans and cooking bananas. It appeared that the ability to modulate the stove's temperature would be a valued feature for cooks.

Our third contribution was measuring the bias caused by observer-induced bias, or the Hawthorne effect. By collecting stove temperature data when technicians were in the home and comparing it to times they were not in the home, we found that households cooked about 2.5 hours per day more on the Envirofit and 2.5 hours less per day on three-stone fires when observers were present and then switched back to previous patterns once the observers had left. We found reductions in wood use (-11.6%) and PM2.5 concentrations (-12.0%) after the introduction of one Envirofit, but once we adjusted for the different behavior when observers were present, this reduction was almost zero. In regard to impacting environmental and health problems, fuel use and particulate matter would need to have declined by much more than what was found in this study. To reach WHO targets for household air pollution, particulate matter needed to decline by 90% from pre-intervention levels. Throughout the study period, three-stone fire use fell by about 2.5 hours a day, but this was more than offset by about 5 hours a day of new cooking on the introduced stoves. This increase in total cooking time diminishes the environmental and household air pollution benefits compared to those shown in the laboratory results. While any reduction in fuel use and particulate matter was likely beneficial for households,[15] fuel-efficient wood stoves such as these will not be adequate to reach safe levels of household air pollution. Thus, policies that assist consumers to shift to safe fuels such as gas or electricity—particularly when coupled with policies to disable smoky indoor stoves—should take on increased importance.

[15] Emerging evidence shows that small reductions in PM2.5 can have benefits in especially vulnerable subpopulations. For example, even a small reduction in PM2.5 can reduce adverse pregnancy outcomes (Alexander et al. 2018) and improve growth in children under the age of two years (LaFave et al. 2019).

Appendix

The details presented here summarize our previous research on how we converted temperature readings into stove usage metrics and measured if the attrition of stove use monitors was random (this appendix is based on Harrell et al. 2016; Simons et al. 2014; 2017; 2018).

A. Placement of SUMs

SUMs must be close enough to the heat source to capture changes in temperatures, but not so close that they exceed 85°C, the maximum temperature the SUMs used in this study can record before they overheat and malfunction. We do not need to recover the exact temperature of the hottest part of the fire to learn about cooking behaviors. Even with SUMs that are reading temperatures 20–30 cm from the center of the fire, as long as the temperature readings for times when stoves are in use are largely different than times when stoves are not used, the logistic regression will be able to predict a probability of usage.

SUMs for three-stone fires were placed in a SUM holder (Figure A3) and then placed under one of the stones in the three-stone fire (left panel, Figure A4). The SUMs for Envirofits were attached using duct tape and wire and placed at the base of the stove behind the intake location for the firewood (right panel, Figure A4). Figure A5 shows an example of SUM temperature data for a household over about three weeks. The left panel shows the temperatures registered in a three-stone fire versus the ambient temperature also recorded with SUMs in this household, while the right panel compares the temperature of the Envirofit to the ambient temperature reading.

B. Visual Observations of Use

Each time data collection personnel visited a household; they observed which stoves were in use (whether the stove was "on" or "off," along with the date and timestamp recorded digitally via a handheld device). Enumerators visited each household several times during a "measurement week," when they also enumerated a survey and weighed wood for the KPT. Another enumerator visited once every 4 to 6 weeks to download data and reset the SUM devices.

C. Generating an Algorithm

We matched the observations of stove use to SUM temperature data by time- and date stamps. At the core of our method was a logistic regression using the lags and leads of the SUM temperature data to predict visual observations of stove usage. We tested 10 specifications of differing combinations of current, lagged, and leading temperature readings (Simons et al. 2014).

In order to determine which of the models was most appropriate, we tested the 10 specifications with the Akaike information criterion (AIC) (Akaike 1981). The AIC trades off goodness of fit of the model with the complexity of the model to guard against over-fitting.

The preferred specification included the temperature reading closest to the time of the observation, the readings 60 and 30 minutes prior, and 60 and 30 minutes after the observation of use, and a control for hour of the day. This regression specification correctly predicted 89.3% of three-stone fire observations and 93.8% of Envirofit observations of stove usage. We then compared our algorithm to other previously published algorithms (Mukhopadhyay et al. 2012; Ruiz-Mercado, Canuz, and Smith 2012). Those algorithms focused on defining "discrete" cooking events based on rapid temperature slope increases and elevated stove temperatures, followed by a cooling off period. We applied those algorithms to the temperature data we had collected and found our logistic regression correctly classified more observations, with a higher pseudo R-squared, than any other algorithm for both three-stone fires and the Envirofits.

D. Random SUMs Attrition

One concern for our study is whether the attrition of the sensors used to measure stove temperatures was random. In cases of sensor malfunction we lost the temperature readings associated with that device (about six weeks of data for that individual stove). The concern is that if damage (overheating above the 85°C tolerance of our SUMs device) was more likely on stoves that were used more heavily, then the data we have are not an unbiased measure of stove usage for the broader sample. If however, the attrition of SUMs sensors is random, there is less concern about the internal validity of our sample.

To examine this topic we follow the approach outlined in Simons et al. (2017) and focus on the endline period where all participants had two three stone fires and two Envirofit stoves. We test this in various ways. First, we regress whether the SUMs data was missing at endline (device malfunctioned) on household fuel wood consumption during that same period. Because fuel wood is a direct input into how much the stoves were used, this is the most direct test of this relationship. If households that cook more (using fuel wood consumption as a proxy) also have a higher probability of SUMs attrition, this would be evidence of non-random attrition and a problem for our study. We examine this relationship separately for each stove type that we included in our study (recall that we choose not to track the non-primary three stone fire by the endline of our study).

Because we are testing for attrition due to excessive cooking (heat exposure) we only test for this relationship on the sample of stoves on which we placed a SUMs device. We also do similar checks with other variables that are related to cooking or experience (count of people cooked for daily, number of meals cooked daily, number of meals in which matooke was cooked daily, and age of the cook).

In Table A1 we present the results of the attrition checks. In our preferred test, we find that the likelihood of SUMs survival is statistically no different than zero (col. 1-3) for each additional kilogram of wood consumption. When examining whether a larger household size is associated with the likelihood of SUMs survival we find a weakly statistically significant relationship for primary three stone fire usage (col. 4). Each additional person cooked for is associated with a four-percentage point decrease in the probability of SUMs survival ($p<0.10$), however this relationship does not appear for either of the Envirofit stoves (col. 5-6). Lastly, we test whether the count of daily meals cooked (col. 7-9), daily meals in which matooke was cooked (col. 10-12), or the age of the cook (col. 13-15) is associated with SUMs survival. We find no statistically significant relationship. Taken as a whole, these tests do not provide strong evidence of non-random attrition of SUMs devices.

Figure 1

Comparison of wood burning stoves: three stone fire versus Envirofit G-3300

(a) Three Stone Fire

(b) Envirofit G-3300

Figure 2

Certified Emissions and Performance Report for Envirofit G3300

April 27, 2011

ENGINES & ENERGY CONVERSION LAB

DEPARTMENT OF MECHANICAL ENGINEERING
COLORADO STATE UNIVERSITY

1374 CAMPUS DELIVERY
FORT COLLINS, CO
80523-1374
970.491.4796
970.491.4799 (F)
WWW.EECL.COLOSTATE.EDU

Emissions and Performance Report

The stove listed below has been tested in accordance with the "*Emissions and Performance Test Protocol*", with emissions measurements based on the biomass stove testing protocol developed by Colorado State University (available at www.eecl.colostate.edu). Percent improvements are calculated from three-stone fire performance data collected at Colorado State University.

Stove Manufacturer:	Envirofit International
Stove Model:	G-3300
Test Dates:	4/4/2011-4/22/2011
Average CO emissions (grams):	18.7
80% Confidence Interval:	17.7-19.7
Percent Improvement:	65.30%
Average PM emissions (milligrams):	995
80% Confidence Interval:	944-1046
Percent Improvement:	51.20%
Average Fuel use (grams):	596.7
80% Confidence Interval:	591.6-601.7
Percent Improvement:	50.10%
Average Thermal efficiency:	32.6
80% Confidence Interval:	32.3-32.8
Percent Improvement:	105.20%
High Power (kW):	3.3
80% Confidence Interval:	3.3-3.4
Low Power (kW):	1.9
80% Confidence Interval:	1.8-1.9

The above results are certified by the Engines and Energy Conversion Laboratory at Colorado State University. All claims beyond the above data are the responsibility of the manufacturer.

Morgan DeFoort
EECL Co-Director
Technical Lead, Biomass Stoves Testing Program

Figure 3
Average Daily Stove Use

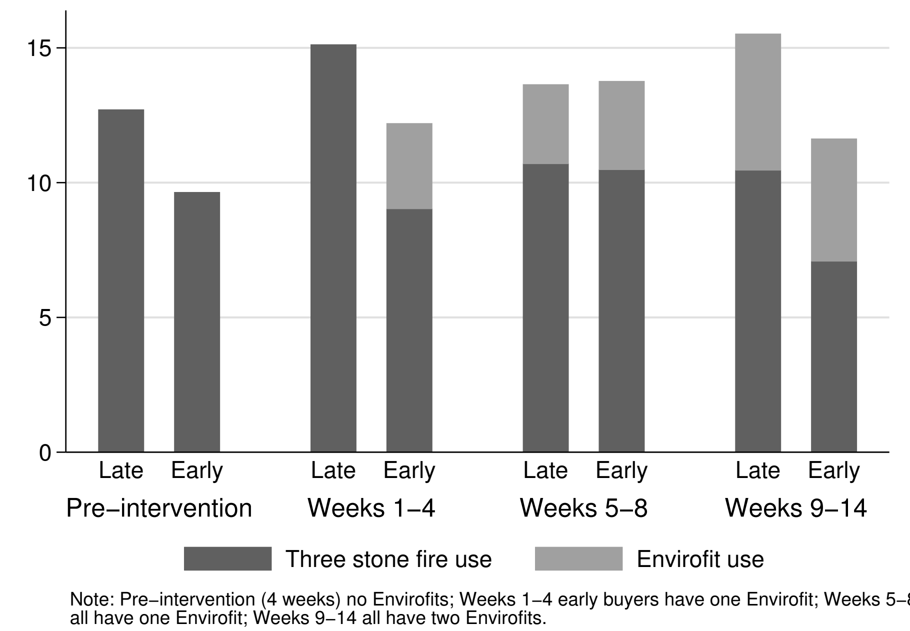


Figure 4
Weekly Stove Use of Early Buyers

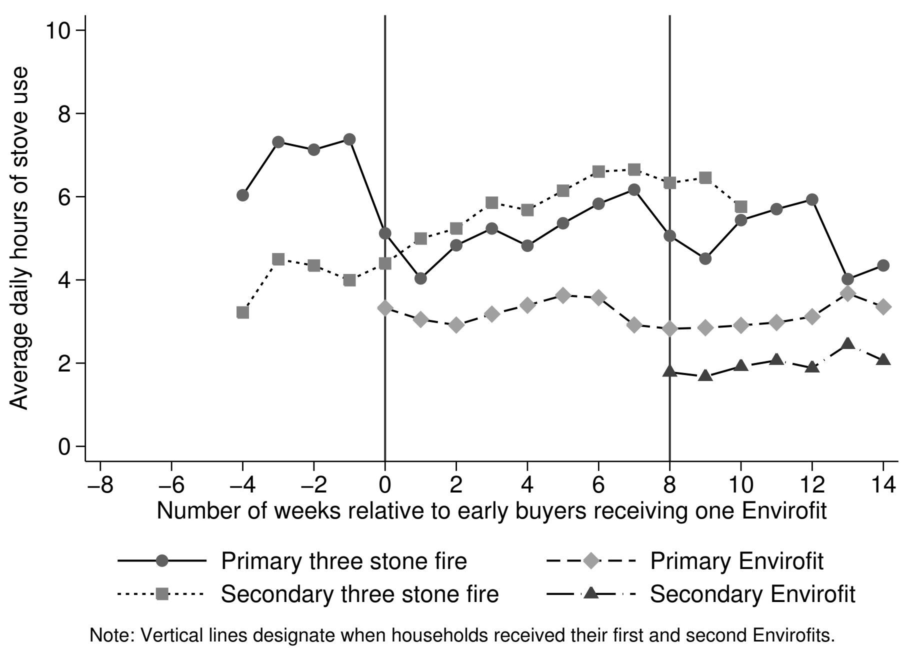


Figure 5
Weekly Stove Use of Late Buyers

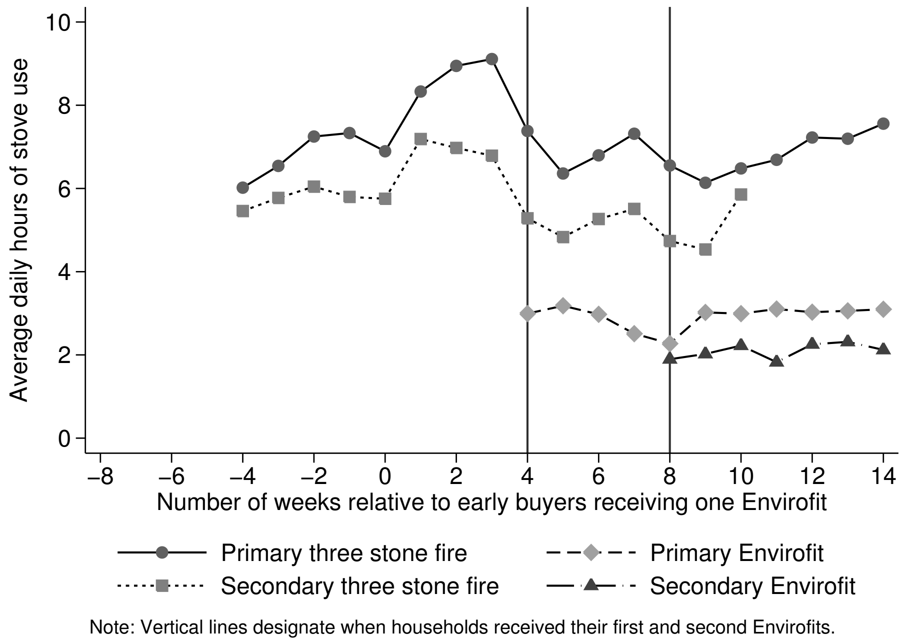


Figure 6
Envirofit Stove with Obvious Signs of Use (from Long Term Usage Study in Nov. 2015)

Table 1
Timeline of Experimental Rollout

Approximate Timing	**Event**
Weeks -4 to week 0	Stove use monitoring (SUMs) begins on two three stone fires
Week 0	Baseline kitchen performance tests (wood weighting) and particulate matter (PM2.5) monitoring*
End of week 0	Deliver first Envirofit to early buyers
Weeks 1-4	SUMs monitoring continues
Week 4	Midline kitchen perfomrance test and PM2.5 monitoring*
End of week 4	Deliver first Envirofit to late buyers (now all participants have one Envirofit)
Weeks 4-8	SUMs monitoring continues
Week 8	Deliver second Envirofit to both early and late buyers
Weeks 8-14	SUMS monitoring continues**
Week 14	Endline kitchen performance test and PM2.5 monitoring*
3.5 years later	Long-term usage follow up

Note: Measurement dates and timing are approximate as roll-out was staggered across the 14 parishes. Stove usage monitors (SUMs) were on all Envirofit stoves and usually on two three stone fires per household.
*Each measurement week (weeks 0, 4, 8) involved three 24-hour periods with wood weighing and particulate matter (PM2.5) monitors.
**After we delivered the second Envirofit stove in week 8 we had a shortage of SUMs, so some homes only had a SUM on one three stone fire.

Table 2

Baseline summary statistics and balance of covariates

	Control Mean	Control SD	Treatment Mean	Treatment SD	Difference	p-value	N
Household demographics							
Female respondent (share)	0.68	0.47	0.73	0.45	0.05	0.38	164
Age of respondent	44.06	13.46	40.38	12.29	-3.68	0.14	163
Married (share)	0.78	0.42	0.77	0.43	-0.01	0.85	164
Wife is primary cook (share)	0.94	0.24	0.92	0.28	-0.02	0.60	164
Spouses make decisions jointly (share)	0.57	0.50	0.52	0.50	-0.05	0.52	164
Socioeconomic status							
Earns income (share)	0.92	0.28	0.88	0.33	-0.04	0.56	163
Self employed (share)	0.73	0.45	0.73	0.45	0.00	1.00	164
Year round employment (share)	0.52	0.50	0.49	0.50	-0.04	0.62	164
Identify as subsistence farmers (share)	0.85	0.36	0.85	0.36	0.00	1.00	164
Value of assets (USD)	905.10	1240.82	1158.37	1650.68	253.27	0.08	164
Stove use and fuels							
Number at largest daily meal	6.16	1.95	6.51	2.25	0.35	0.23	163
Always boils drinking water (share)	0.74	0.44	0.72	0.45	-0.02	0.69	164
Firewood primary fuel source (share)	0.94	0.24	0.95	0.22	0.01	0.81	164
Purchased firewood last month (share)	0.34	0.48	0.43	0.50	0.08	0.24	162
Gathered firewood last month (share)	0.82	0.39	0.81	0.39	-0.01	0.97	163
Baseline cooking measurements							
Daily hours cooked on primary three stone fire	7.30	6.75	8.14	7.21	0.84	0.47	118
Daily hours cooked on secondary three stone fire	5.91	6.41	4.51	5.68	-1.41	0.28	99
Daily hours cooked on all three stone fires	12.43	9.71	10.34	8.99	-2.10	0.34	91
Net wood used daily (weight in kg)	9.30	4.10	10.02	4.70	0.73	0.38	153
Average PM2.5 reading, $\mu g/m^3$	414.30	240.84	372.66	228.91	-41.64	0.33	150
Number of households receiving offer	82		82				164

Note: Household data collected at parish wide sales meetings. We adjust standard errors for clustering at the parish level. To minimize the effect of outliers the value of assets and PM2.5 readings are top and bottom coded at 98% and 2% of the distribution while wood use is top coded at 98%. Daily hours cooked on all three stone fires is only calculated if non missing values exist for both the primary and secondary three stone fire. The prices used to calculate asset values are taken from the 2011-12 round of the Uganda Living Standards Measurement Survey (LSMS) published by the World Bank. Values presented are rounded to two decimal places, the value in the difference column is calculated prior to rounding.

Table 3

Envirofit stove use

Variable	Mean	Std. Dev.	Min.	Max.	N
Early buyers have one Envirofit					
Daily hours cooked on primary Envirofit	4.35	3.89	0.02	16.75	188
All buyers have two Envirofits					
Daily hours cooked on primary Envirofit	4.25	3.68	0	16.23	198
Daily hours cooked on secondary Envirofit	2.91	3.5	0	16.93	198
Daily hours cooked on all Envirofits	7.17	4.79	0.26	24.59	198

Note: This table only includes data from weeks with a kitchen performance test when households had one or two Envirofits.

Table 4

Effect of the Envirofit on daily wood used for cooking

Dependent variable = kg. of wood used daily

VARIABLES	(1) OLS	(2) FE
Treatment	0.72	
	(0.72)	
Early buyers have one Envirofit	-1.86***	-1.73***
	(0.60)	(0.56)
All buyers have two Envirofits	-2.48***	-2.48***
	(0.68)	(0.66)
Treatment x Early buyers have one Envirofit	-0.95	-1.08*
	(0.85)	(0.56)
Treatment x All buyers have two Envirofits	-0.46	-0.55
	(0.88)	(0.59)
Constant	12.40***	
	(0.46)	
Observations	1,116	1,116
R-squared	0.15	0.42
Number of household fixed effects		163

Standard errors clustered at parish-wave level in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Wood weights are top coded at 98%. OLS regressions include parish fixed effects.

Table 5

Effect of the Envirofit on daily PM concentrations

Dependent variable = natural log daily PM concentrations		
	(1)	(2)
VARIABLES	OLS	FE
Treatment	-0.02	
	(0.03)	
Early buyers have one Envirofit	0.12**	0.12**
	(0.05)	(0.05)
All buyers have two Envirofits	-0.10**	-0.10*
	(0.04)	(0.05)
Treatment x Early buyers have one Envirofit	-0.13*	-0.12**
	(0.07)	(0.06)
Treatment x All buyers have two Envirofits	-0.02	-0.02
	(0.06)	(0.06)
Constant	6.57***	
	(0.07)	
Observations	1,242	1,242
R-squared	0.87	0.92
Number of household fixed effects		164

Standard errors clustered at parish-wave level in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: OLS regression includes parish fixed effects and all regressions include PM monitor fixed effects. PM2.5 readings are top and bottom coded at 98% and 2% of the distribution prior to taking the natural log.

Table 6

Effect of the Envirofit on daily hours cooked on three stone fires - all weeks

Dependent variable = daily hours cooked on all (cols. 1 and 2),
primary (cols. 3 and 4), or secondary (cols. 5 and 6) three stone fire(s)

VARIABLES	(1) OLS	(2) FE	(3) OLS	(4) FE	(5) OLS	(6) FE
Treatment	-2.58		0.26		-1.86	
	(2.47)		(1.36)		(1.22)	
Weeks 1-4 (Early buyers have one Envirofit)	1.80	1.96**	1.28	1.49*	0.82	1.22***
	(1.79)	(0.83)	(1.00)	(0.84)	(0.82)	(0.32)
Weeks 5-8 (All buyers have one Envirofit)	-2.72	-3.09***	0.34	0.42	-0.73	-1.04**
	(1.82)	(0.95)	(1.19)	(0.88)	(0.90)	(0.42)
Weeks 9-14 (All buyers have two Envirofits)	-3.61*	-5.15***	-0.45	-0.38	-0.13	-0.85
	(2.08)	(1.53)	(1.15)	(0.91)	(0.94)	(0.62)
Treatment x Early buyers have one Envirofit	-3.16	-3.73***	-3.33**	-3.68***	0.15	-0.58
	(2.67)	(0.74)	(1.60)	(1.12)	(1.37)	(0.48)
Treatment x All buyers have one Envirofit	1.83	2.89***	-1.91	-1.77	2.96**	3.07***
	(2.78)	(1.05)	(1.86)	(1.09)	(1.35)	(0.78)
Treatment x All buyers have two Envirofits	-0.29	0.73	-1.47	-1.03	2.66	1.19
	(3.18)	(1.75)	(1.96)	(1.25)	(1.68)	(1.07)
Constant	14.39***		5.63***		6.27***	
	(1.76)		(0.92)		(0.92)	
Observations	8,595	8,595	13,890	13,890	8,056	8,056
R-squared	0.13	0.58	0.10	0.45	0.08	0.52
Number of household fixed effects		144		160		146

Standard errors clustered at parish-wave level in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Data includes all weeks that temperature sensors were on stoves. OLS regressions include parish fixed effects.

Table 7

Effect of the Envirofit on daily hours cooked on three stone fires - measurement weeks

Dependent variable = daily hours cooked on all (cols. 1 and 2), primary (cols. 3 and 4), or secondary (cols. 5 and 6) three stone fire(s)

VARIABLES	(1) OLS	(2) FE	(3) OLS	(4) FE	(5) OLS	(6) FE
Treatment	-1.93		0.78		-0.91	
	(2.00)		(1.01)		(1.14)	
Early buyers have one Envirofit	4.35**	2.75	2.56**	3.77***	2.17**	1.55
	(1.93)	(1.95)	(1.16)	(1.01)	(0.86)	(0.94)
All buyers have two Envirofits	-3.56	-10.20**	-1.49	-0.86	1.06	0.94
	(2.85)	(3.81)	(1.19)	(1.34)	(1.62)	(2.40)
Treatment x Early buyers have one Envirofit	-7.41***	-6.36***	-6.56***	-7.79***	-1.09	-1.07
	(2.52)	(1.63)	(1.57)	(1.17)	(1.49)	(0.99)
Treatment x All buyers have two Envirofits	-3.16	3.38	-2.42	-2.53	1.71	0.30
	(3.71)	(4.71)	(1.83)	(1.74)	(3.38)	(3.75)
Constant	12.36***		5.06***		6.73***	
	(1.62)		(0.94)		(0.79)	
Observations	571	571	941	941	555	555
R-squared	0.24	0.73	0.18	0.60	0.13	0.73
Number of household fixed effects		129		155		133

Standard errors clustered at parish-wave level in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: This table only includes data from weeks with a kitchen performance test. OLS regressions include parish fixed effects.

Table 8

Effect of the Envirofit on daily hours cooked on Envirofit(s) - all weeks

Dependent variable = daily hours cooked on all (cols. 1 and 2), primary (cols. 3 and 4), or secondary (col. 5) Envirofit(s)

VARIABLES	(1) OLS	(2) FE	(3) OLS	(4) FE	(5) OLS
Treatment	0.44		0.44		0.07
	(0.35)		(0.55)		(0.35)
Weeks 5-8 (All buyers have one Envirofit)	-0.17	0.05	-0.09	-0.02	
	(0.27)	(0.26)	(0.25)	(0.21)	
Weeks 9-14 (All buyers have two Envirofits)	1.90***	2.24***	0.08	0.04	
	(0.56)	(0.54)	(0.50)	(0.33)	
Treatment x All buyers have two Envirofits	-0.76	-0.90	-0.22	-0.22	
	(0.72)	(0.56)	(0.51)	(0.29)	
Constant	1.59***		1.53**		2.16***
	(0.43)		(0.64)		(0.10)
Observations	6,853	6,853	8,923	8,923	2,957
R-squared	0.12	0.47	0.09	0.41	0.10
Number of household fixed effects		130		152	

Standard errors clustered at parish-wave level in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Data includes all weeks that temperature sensors were on stoves. OLS regressions include parish fixed effects. The constant in column (1) corresponds to the period when early buyers owned one Envirofit.

Table 9

Effect of the Envirofit on daily hours cooked on Envirofit(s) - measurement weeks

Dependent variable = daily hours cooked on all (cols. 1 and 2), primary (cols. 3 and 4), or secondary (col. 5) Envirofit(s)

VARIABLES	(1) OLS	(2) FE	(3) OLS	(4) FE	(5) OLS
Treatment	-0.01		0.21		0.65
	(0.77)		(0.66)		(0.48)
All buyers have two Envirofits	2.71***	3.08***	0.10	-0.36	
	(0.65)	(0.81)	(0.54)	(0.57)	
Constant	3.97***		3.75***		3.00***
	(0.77)		(0.66)		(0.14)
Observations	390	390	482	482	256
R-squared	0.16	0.66	0.05	0.57	0.12
Number of household fixed effects		105		129	

Standard errors clustered at parish-wave level in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: This table only includes data from weeks with a kitchen performance tests. At midline treatment households owned one Envirofit and at endline all households owned two Envirofits. OLS regressions include parish fixed effects. The constant in column (1) corresponds to the period when early buyers owned one Envirofit.

Table 10
Adjustments for Hawthorne effect

	Change in TSF Hours (hr/day)	TSF wood usage (kg/hr)	Change in ENV Hours (hr/day)	ENV wood usage (kg/hr)	Adjustment for Wood (kg/day)
Midline (Early Buyers)	2.63	0.64	-2.38	0.32	0.92
Endline (All Buyers)	5.05	0.64	-3.06	0.32	2.25

	Change in TSF Hours (hr/day)	TSF PM2.5 (μg/m^3 per hr)	Change in ENV Hours (hr/day)	ENV PM2.5 (μg/m^3 per hr)	Adjustment for PM2.5 (μg/m^3 per day
Midline (Early Buyers)	2.63	32.95	-2.38	16.08	48.39
Endline (All Buyers)	5.05	32.95	-3.06	16.08	117.19

Note: Stove users used three stone fires less and Envirofit stoves more when observers were present, when observers departed they reveresed these changes (Simons et al. (2017)). Therefore, to adjust for this observer induced bechavior, we calculate the change in TSF hours per day as the difference in the coefficients when estimating the effect of the introduction of Envirofit(s) on TSF use only in the measurement week compared to all weeks (difference of coefficients between Table 6 and 7). The change in ENV hours per day is calculated as the difference in the coefficients when estimating the effect of the introduction of Enviroift(s) on ENV use only in the measurement week compared to all week (difference of coefficients between Table 8 and 9). Three stone fire wood (PM2.5) usage per hour calculated during first kitchen performance test when no one owned an Envirofit. Envirofit wood (PM2.5) usage per hour calculated using the laboratory results shown in the "Emission and Performance Report" (Figure 2) because we do not have any periods in our experimental setting when households only had Envirofits.

Table 11

Estimates of Wood Use and PM concentrations after Hawthorne Effect Adjustment

	Baseline Amount (kg/day)	Unadjusted Change (kg/day)	Unadjusted Change (%)	Adjustment (kg/day)	Adjusted Change (kg/day)	Adjusted Change (%)
Midline (Early Buyers)	9.30	-1.08	-11.6%	0.92	-0.16	-1.7%
Endline (All Buyers)	9.30	-2.48	-26.7%	2.25	0.48	-2.5%

	Baseline Amount (μg/m^3 per day)	Unadjusted Change (μg/m^3 per day)	Unadjusted Change (%)	Adjustment (μg/m^3 per day)	Adjusted Change (μg/m^3 per day)	Adjusted Change (%)
Midline (Early Buyers)	414.30	-49.72	-12.0%	48.39	-1.33	-0.33%
Endline (All Buyers)	414.30	-41.43	-10.0%	117.19	75.76	18.3%

Note: Unadjusted estimates of the change in wood usage come from Table 4. Unadjusted estimates of the change in PM2.5 come from Table 5. The adjustments are calculated in Table 10. Calculations for the adjusted changes are based on Equations 3 and 4. Baseline amounts come from Table 2.

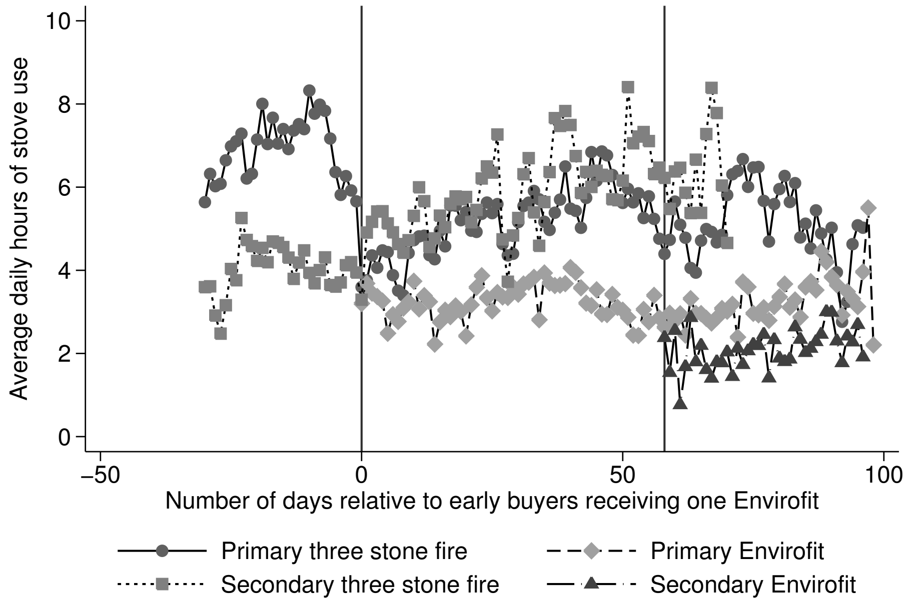
Table 12

Long term usage study: unannounced home visit 3.5 years after initial Envirofits delivered

	N	%
Someone home for unannounced long term usage study	137	100.0%
Actively cooking in moment when enumerators arrived	66	100.0%
-among those, cooking on three stone fire only	52	78.8%
-among those, cooking on Envirofit only	6	9.1%
-among those, cooking on other (mud/charcoal) stove	8	12.1%
Among all households not using Envirofit when enumerators arrived, enumerators asked to see primary Envirofit stove for signs of use	131	100.0%
-primary Envirofit with obvious signs of use	85	64.9%
-primary Envirofit stored and clearly not being used	22	16.8%
-primary Envirofit stored and in perfect condition (basically never used)	3	2.3%
-primary Envirofit damaged and disposed of	11	8.4%
-primary Envirofit given away (condition unknown)	10	7.6%
Among all households that stated they received two Envirofits, enumerators asked to see secondary Envirofit stove for signs of use	129	100.0%
-secondary Envirofit with obvious signs of use	32	24.8%
-secondary Envirofit stored and clearly not being used	14	10.9%
-secondary Envirofit stored and in perfect condition (basically never used)	12	9.3%
-secondary Envirofit damaged and disposed of	21	16.3%
-secondary Envirofit given away (condition unknown)	49	38.0%
Asked: "Do you still use the Envirofit stove?"	137	100.0%
-"I still use both Envirofits"	31	22.6%
-"I still use only one Envirofit"	69	50.4%
-"I have stopped using Envirofits"	37	27.0%
Asked: "If you bought a new stove today, would you purchase an Envirofit?"	137	100.0%
-Yes	108	78.8%
-No	21	15.3%
-Unsure or no response	8	5.8%

Figure A1

Daily stove use of early buyers



Note: Vertical lines designate when households received their first and second Envirofits.

Figure A2

Daily stove use of late buyers

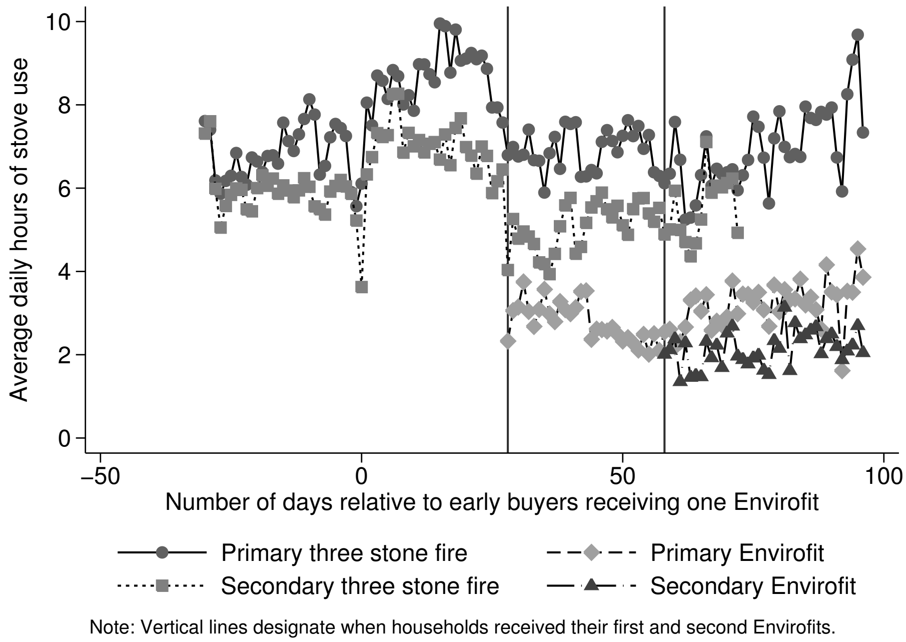


Figure A3

SUM holder designed to encase the stove use monitor to protect it from malfunctions when exceeding temperatures of 85 degrees Celsius

Figure A4

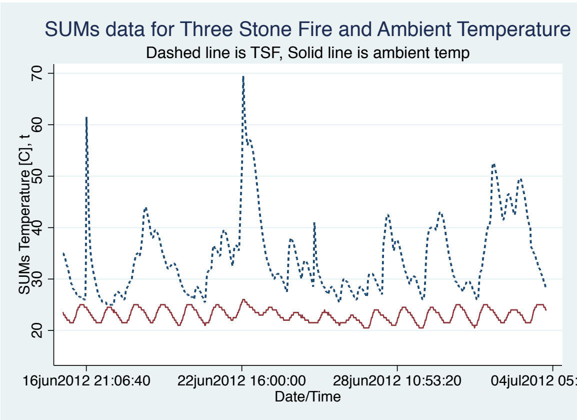
Arrows mark the placement of SUMs on three stone fire and Envirofit

(a) Three Stone Fire

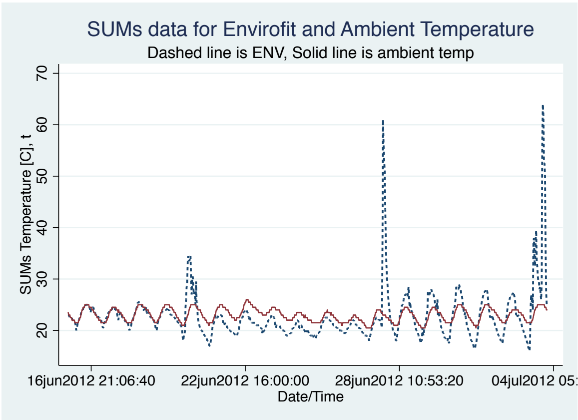
(b) Envirofit

Figure A5

Example of household level SUMs temperature data in same household at same times



(a) Three Stone Fire



(b) Envirofit

Table A1

SUMs Device Attrition: linear probability model of missing SUMs data

	TSF1 (1)	ENV1 (2)	ENV2 (3)	TSF1 (4)	ENV1 (5)	ENV2 (6)	TSF1 (7)	ENV1 (8)	ENV2 (9)	TSF1 (10)	ENV1 (11)	ENV2 (12)	TSF1 (13)	ENV1 (14)	ENV2 (15)
Daily household wood use (kg)	-0.02	-0.01	-0.01												
	(0.01)	(0.02)	(0.01)												
People cooked for daily				-0.04*	-0.00	-0.00									
				(0.02)	(0.02)	(0.02)									
Meals cooked daily							-0.09	-0.07	0.04						
							(0.06)	(0.05)	(0.06)						
Meals cooked *matooke* daily										0.02	0.10	0.09			
										(0.09)	(0.08)	(0.13)			
Age of cook													-0.00	0.00	-0.00
													(0.00)	(0.00)	(0.00)
Observations	147	142	137	153	146	141	153	146	141	153	146	141	153	147	141
R-squared	0.02	0.00	0.01	0.03	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.00
Parish clusters	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14

Standard errors clustered at parish level in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: The dependant variable is a 0/1 for whether we have the SUMs temperature data during the endline for a specified stove type (TSF1 = primary three stone fire, ENV1 = first Envirofit, ENV2 = second Envirofit). The overall sample only includes stoves that we placed a SUMs device on during the endline. The daily wood weights and counts of cooking practices are averaged across the KPT measurement week. To account for possible correlation in how data was collected by the measurement team, we cluster standard errors at the parish level because the KPT measurement team spent a week at a time in a given parish.

Chapter 3

What is a "Meal"? Comparative Methods of Auditing Carbon Offset Compliance for Fuel-Efficient Cookstoves

I. Introduction

Globally, approximately 2.8 billion people cook on traditional stoves that burn solid fuels such as wood and charcoal (Bonjour et al. 2013). The burning of solid fuels associated with traditional cookstoves, such as the three-stone fire, has high global health costs, causing an estimated four million deaths per year (Lim *et al.* 2012), and high costs in terms of the release of greenhouse gases (Robert Bailis, Ezzati, and Kammen 2005) and black carbon (soot), which contribute to global warming (Bond, Venkataraman, and Masera 2004; Ramanathan and Carmichael 2008). Inefficient stoves also contribute to deforestation (Arnold, Köhlin, and Persson 2006).

Because fuel-efficient cookstoves (defined as stoves that use less fuel than the three-stone fire or relevant baseline stove) release less carbon dioxide than traditional three-stone fires, some fuel-efficient cookstove projects have received carbon credits that subsidize the cost of the stoves (Simon, Bumpus, and Mann 2012). Fuel-efficient cookstoves are an especially attractive target for carbon credits because, in addition to decreases in greenhouse gas emissions, these stoves can also improve the health and safety of users and reduce the time people (usually women and children) spend collecting fuel (Burke and Dundas 2015; Kammen, Bailis, and Herzog 2002).

To quantify the changes in cooking practices due to fuel-efficient cookstoves, it is not sufficient to simply measure the usage of the new stove. It is also crucial to measure any reduction in the use of traditional stoves (Ruiz-Mercado *et al.* 2011; Miller and Mobarak 2013) as many owners of new stoves continue to use old stoves and fuels in a phenomenon known as "stove stacking" (Masera, Saatkamp, and Kammen 2000; Ruiz-Mercado *et al.* 2011), which reduces the environmental benefits of using only the fuel-efficient cookstoves. Researchers examining the effects of cookstove programs need to understand and measure these nuances to correctly estimate program benefits. The carbon market needs to know whether these new stoves really do offset the amount of greenhouse gases that they claim to reduce. The Gold Standard methodology calculates carbon credits for cookstove projects based on reductions in carbon emissions estimated through reductions in fuel use after a fuel-efficient stove is purchased.[1] To estimate the reduction in carbon emissions, the project developer estimates biomass fuel savings, the fraction of non-renewable biomass, and emission factors for fuel consumption (Lee et al. 2013). These measures feed into an equation with default emission factors and global warming potentials to derive the CO_2-equivalent saved for a given project year. An auditor comes to the project site to spot check data, calculations, and visits a small number of households. If there are errors, problems with study design, or the auditors' observations deviate from what was

[1] For more details, see: http://www.cdmgoldstandard.org/frequently-asked-questions/carbon-market.

reported by the project developer, then the auditor may recommend that the offsets be adjusted or additional data be collected.

This paper focuses on understanding stove use in a variety of ways. We compare how different measures of cooking correlate across 163 households (403 household-days) in Mbarara, Uganda, all of whom used three-stone fires as their normal cooking technology. We measured hours cooked (derived from a predictive logistic regression of stove usage monitors on observations of stoves in use), detailed household food diaries, weight of wood consumed, and household air pollution ($PM_{2.5}$, particulate matter less than 2.5 micrometers in diameter) concentrations. We do these analyses both across and within households.

Our goal is to identify how well these measures predict each other. If some measures correlate strongly with others, then it is possible that carbon offset auditors and/or researchers with intensive monitoring programs can rely on a subset of measures (or a single measure). If an inexpensive method's results strongly predict a more costly method's results, perhaps that method can be used, reducing overall monitoring costs. Using a proxy could be particularly useful in the case of indoor air pollution monitoring because the equipment required to measure particulate matter is costly and requires technical oversight. Conversely, if one measure appears unrelated to the other measures, then it may not be valid and should not be used in isolation. If all the measures correlate positively but all the correlations are weak, then carbon compliance officers and related researchers must continue to improve measurement techniques. Without more robustly correlated measures, multiple measures will be necessary to create confidence in stove use metrics, their related quantity of carbon offsets, and their related use in the measurement of other benefits of fuel-efficient cookstove programs, such as indoor air pollution reductions and time savings.

The paper is organized as follows. Section II describes the study area and the previous methodological studies that compared techniques to measure stove use. Section III describes the research methods used. Section IV reports the results and discusses their importance. Lastly, section V concludes and discusses the policy implications for researchers and auditors of fuel-efficient cookstove programs.

II. Measuring Stove Usage

We analyze data from daily household visits over four days, yielding three 24-hour periods of measurement. During each 24-hour period of measurement, we recorded temperatures on each three-stone fire every 30 minutes using stove usage monitors (SUMs), which we regress on visual observations of stoves in use to determine predicted hours cooked; food diaries consisting of foods cooked, type(s) of fuel(s) used, type(s) of stove(s) used, number of stoves used, and number of people cooked for at each meal; the amount of fuel used via kitchen performance tests (KPTs); and mean 24-hour particulate matter concentrations of $PM_{2.5}$ using University of California, Berkeley Particle and

Temperature Sensors (UCB-PATS). Because all households used three-stone fires, these measurement techniques capture near complete household-level data.[2]

A. Prior Studies Comparing Measures of Stove Usage

A few studies have compared predicted time spent cooking (based on an algorithm using stove temperature data) with another method for measuring stove usage. These past studies have used two methods to determine fuel-efficient stove usage and impact. For example, Thomas *et al.* (2013) compared predicted time spent cooking with self-reported use and direct observations. Ruiz-Mercado *et al.* (2013) compared predicted time spent cooking with cooking behaviors described in food diaries. Pillarisetti *et al.* (2014) used predicted time spent cooking over more than a year to determine whether use early in the year predicted long-term use. Graham *et al.* (2014) used predicted time spent cooking to estimate fuel consumption and then compared the estimates with observations in the field.

Our study expands upon these previous studies by directly comparing a number of stove measurement techniques in the same setting, making the unique contribution of comparing four methods of stove measurement.

B. Study Area

For our study site, we selected the Mbarara region of Uganda because it is rural, almost all families cooked on a traditional three-stone fire, there was no active fuel-efficient cookstove intervention in the region, it was less than a day's travel from Kampala, and families spent a lot of time gathering wood (approximately 10-20 hours per week).[3] The main economic activity is agrarian, including farming of *matooke* (a type of green banana), potatoes, and millet, as well as raising livestock. We executed a series of randomized control trails surrounding the drivers of purchase and usage of fuel-efficient cookstoves. The present study highlights the measurement techniques we employed for measuring stove usage. However, we also studied the impact of informational marketing messages and liquidity constraints on the purchase decision (Beltramo et al. 2015b; Levine et al. 2018), as well as social network effects on purchase (Beltramo *et al.* 2015a). Those studies can be consulted for a more detailed background on the Mbarara region.

[2] All households used at least one three stone fire and typical practice was two three stone fires per household. Across the entire sample of cookstoves observed 97% were three stone fires, 2% mud stoves, and 1% charcoal stoves.

[3] Wood was scarcer in some northern parts of Uganda, but given the poor road infrastructure in those districts, they were logistically too difficult for us to work in.

III. Methods

A. Stove Measurement Usage Techniques

Our first usage measure is estimated hours cooked generated from continuously recorded stove temperature data and visual observations of use. The temperature data and visual observations are processed with an algorithm based on Simons *et al.* (2014a) to determine an estimate of the minutes cooked on each stove in a household. Throughout the experiment, field staff recorded visual observations (about 2,400 visual observations) of whether a stove was in use (on/off)[4] when they visited homes. Then we used a machine-learning algorithm to examine the temperature data immediately before and after the 2,400 visual observations of use. The algorithm analyzed the data to understand how temperature patterns change at times of observed stove use and then predicted cooking behaviors to the wider dataset of temperature readings.[5]

The goal of researchers examining environmental benefits (and for auditing for carbon offsets) is to compare the carbon released by households before and after the introduction of a fuel-efficient stove, the development of algorithms for temperature data for three-stone fires and fuel-efficient stove types could allow for longer periods of comparison (i.e., months) compared to kitchen performance tests that measure kilograms of wood used, which usually last a week or less. Using SUMs offer an unobtrusive, relatively inexpensive, and objective measure of stove usage, however, a concern with SUMs is that they record temperature, and not exact stove usage. A companion paper (Simons *et al.* 2014a), discusses many of the slippages between temperature and cooking.

We used iButtons**™** as temperature logging devices for our SUMs**,** which are small stainless steel temperature sensors about the size of a small coin and the thickness of a watch battery that can be affixed to a stove or open fire and record temperatures with an accuracy of +/- 1.3°C up to 85°C.[6] The iButtons were set to record temperatures every 30 minutes.[7] We buried one iButton below each three-stone fire, as three-stone fires comprised 97% of stoves used in our study area.[8] Typically, households have one larger three-stone fire to cook the main part of the meal (usually *matooke* and/or beans) and a smaller three-stone fire to cook side dishes and sauces. In these instances, we used one

---

4 Enumerators were instructed to mark whether a household was cooking (on/off) based on the presence of a flame or hot coals and food being cooked each time they entered a participant home.

5 For this paper we modified the Simons *et al.* (2014a) algorithm by adjusting the probability of nighttime cooking to be zero if the predicted probability of cooking is less than 0.85 during nighttime hours. Because enumerators did not visit homes during the night (too intrusive for participant households) we have no visual observations of stove use taken during the night, therefore this adjustment is necessary. The Simons *et al.* (2014a) study only examines how well the algorithm predicts visual observations of stove use (all taken during the day) and therefore does not deal with potential adjustments that are necessary to apply the technique to nighttime temperature readings.

6 For additional details concerning the iButtons, see: http://www.berkeleyair.com/products-and-services/instrument-services/78-sums.

7 A shorter period between temperature readings (such as 10 or 20 minutes) would have been in line with other studies (Pillarisetti et al. 2014; Ruiz-Mercado, Canuz, and Smith 2012), however due to logistical and budgetary constraints (tracking every 10 minutes would have required a technician to download data and reset the device every two weeks, tracking every 20 minutes would require a visit once every four weeks) we opted for a 30 minute interval which requires a technician visit only once every six weeks. Due to the size and duration of our study, and a desire to minimize household visits so as not to influence participant behavior, a shorter resolution than 30 minutes was not possible. In Simons et al. (2014a) using data from this same setting and 30 minute resolution we correctly predicted 89% of visual observations of use in the three stone fires.

8 Due to the extremely small sample sizes of mud stoves (2%) and charcoal stoves (1%) in the study area, the analysis only covers three-stone fires.

iButton for each three-stone fire. Approximately every six weeks, we collected the iButtons, downloaded the temperature data, and replaced them with new iButtons. For a more detailed description of the data collection process with iButtons, see Simons *et al.* (2014a).

Our second measurement technique is through the use of food diaries (Prentice 2003; Krall and Dwyer 1987). While food diaries create a detailed account of everything cooked in a given household, they can be inaccurate due to recall bias and experimenter demand effects (if respondents over-report the use of the item, such as stoves or foods, that the experimenter is interested in). Food diaries also do not directly measure the duration of cooking. Another potential complexity with self-reported food diaries is households cooking two meals at once but then reporting the two meals as separate events.

Our third measure is a KPT, which weights the woodpile in a kitchen on sequential days to quantify the amount of wood used in a given 24-hour period.[9] The KPT is the protocol used to estimate fuel savings, a primary component of calculating carbon credits for a stove project (The Gold Standard Foundation 2013). To minimize variance, the standard recommendation is that the KPT testing period should be at least three days, avoiding weekends and holidays (Bailis, Smith, and Edwards 2007)[10]. Although the KPT is a useful tool for measuring fuel consumption, there are challenges in carrying out the protocol. Changes in the weight of a wood pile may not equal wood burned due to households sharing wood with neighbors, households inadvertently adding wood to the measured pile of wood, or wood becoming wet or dry between initial and final weighing. Additionally, direct observational processes alter participants' behavior (as noted by Ezzati, Saleh, and Kammen 2000; Smith-Sivertsen *et al.* 2009; Simons *et al.* 2017).

Our fourth measurement technique is using PM monitors to measure the mean 24-hour concentrations of particulate matter (PM). PM monitors measure the concentration of particles in wood smoke that have negative health effects (McCracken *et al.* 2007; Smith *et al.* 2010). However, to our knowledge, PM concentrations have not been used to estimate stove usage because PM concentration potentially depends on stove type, fuel, cooking practice (high or low temperature, smoldering wood, etc.), airflow in the kitchen, moisture content of wood, proximity of cook to the fire, ambient background PM levels, and other factors. We measured mean 24-hour concentrations of $PM_{2.5}$ by installing calibrated UCB-PATS PM monitors in the study participants' homes during the same 72 hours of the kitchen performance test. We followed best practices as outlined by the Berkeley Air Monitoring Group[11] and measured three consecutive days of mean 24-hour $PM_{2.5}$ concentrations in the kitchen. We averaged data from the UCB-PATS PM monitors into 24-hour average $PM_{2.5}$ readings in $\mu g/m^3$.

B. Statistical Methods

We compared each of the cooking event measurements using ordinary least squares pooled across days and household, instrumental variables, and within-household (fixed-effects) regressions. Pooled regressions were clustered by household. Instrumental variables allow for consistent estimation in cases where covariates are measured with measurement error (which is possible in our setting). Within-

[9] For more information, see: http://ehs.sph.berkeley.edu/hem/content/KPT_Version_3.0_Jan2007a.pdf.
[10] For more information, see: http://cleancookstoves.org/technology-and-fuels/testing/protocols.html.
[11] For details, see: http://berkeleyair.com/services/ucb-particle-and-temperature-sensor-ucb-pats/.

household estimators eliminate the bias of time-invariant omitted variables by including a household fixed-effect. However, because the within-household estimations only consider changes in covariates over time, our identifying variation is very limited and likely attenuates our point estimates downward. We ran ordinary least squares (OLS) regressions with the following specification:

$$Y_{it} = \beta_0 + \beta_j X_{jit} + u_{it},$$

where Y_{it} is a cooking measure (time spent cooking, wood use, or PM) at household i on day t, X_{jit} are alternative cooking measures j (number of meals cooked, cooking *matooke* or beans, number of people cooked for, etc.) at household i on day t, β_j is the coefficient estimate associated with cooking measure j, and u_{it} is a residual for that household on that day.

We used an instrumental variables approach to allow for consistent estimates in cases where the covariates are measured with error, but the error terms are orthogonal.[12] As time spent cooking, wood use, and particulate matter were all measured with error, we instrumented for each of these using food diary data (number of meals cooked per day, number of instances of cooking beans or *matooke* per day, and the maximum number of people cooked for at lunch or dinner). We used two-stage least squares, and in the first stage regress the covariate measured with error on the three food diary instruments ($\boldsymbol{Z}$),

$$X_{jit} = \delta' \boldsymbol{Z} + u_{it}.$$

Then we estimated the predicted value of this regression, $\hat{X} = \hat{\delta}' \boldsymbol{Z}$. In the second stage, the regression of interest was estimated as usual, except the covariate measured with error was replaced with the predicted values from the first stage, $Y_{it} = \hat{X}\beta_{iv} + \epsilon_{it}$. In cases where the first stage was not estimated with high precision, we did not proceed to the second stage.

We also ran regressions with a fixed effect for each household (v_i):

$$Y_{it} = \beta_0 + \beta_j X_{jit} + v_i + \varepsilon_{it}.$$

The coefficient β_j is the estimate of how increasing one measure of stove usage (j) at a household predicts higher levels of another measure of stove usage at that household on a different day of that week. The fixed effects estimator has the advantage that fixed attributes of the home that week (e.g., ventilation or placement of the iButton) do not affect the estimate. The disadvantage is that it relies only on within-household variation across three or so measurement days. Thus, its precision is low if households do not change their measured behavior much across adjacent days.

[12] This use of instrumental variables is not related to achieving causal estimates, but merely addresses measurement error. Our results remain descriptive about how different measures correlate.

IV. Results and Discussion

A. Common Cooking Practices and Descriptive Statistics

There are four main meals cooked in the Mbarara region: breakfast, lunch, afternoon tea, and dinner. Almost all families cook on a traditional three-stone fire, usually located within a cooking hut. In our sample, 62% of households had no windows in the cooking hut. Most stove usage occurs during lunch and dinner preparation, with *matooke* and beans as the most common and most time-consuming foods cooked. *Matooke*, the main food for lunch and dinner, is typically steamed for three to five hours. Beans are prepared by boiling and simmering for two to four hours. Households in this region use stoves exclusively for cooking, not for heating or other purposes.

Table 1 includes summary statistics for hours cooked based on the predicted logit specification for the selected days of the kitchen performance tests. These statistics correspond to the predicted stove usage for the 219 24-hour periods when we also had wood weighing, food diaries, and PM monitors. Of these days, the main three-stone fire was used on average for approximately 6 hours and 26 minutes and the secondary three-stone fire was used on average for approximately 4 hours and 31 minutes.

Respondents reported cooking an average of 3.34 meals per day (Table 1). Most stove usage occurs during the preparation of the two largest meals, lunch and dinner; thus, the analysis focuses on these two meals. The average value of the maximum number of people at either lunch or dinner, which is our main measurement from the food diaries, was 6.34. Snack/tea had the lowest average number of attendees (4.4), while dinner had the largest (6.1). For lunch, *matooke* was cooked on 78% of days and beans were cooked on 42% of days. For dinner, *matooke* was cooked on 71% of days and beans were cooked on 56% of days.

There are 359 measures of daily wood weights.[13] Mean daily wood use was 9.91 kilograms. After top-coding the highest 5%, the mean amount of wood used in a 24-hour period was 9.62 kilograms (Table 1).

We measured 365 days of particulate matter concentrations, with a mean 24-hour PM concentration of 1019 $\mu g/m^3$ (Table 1). This level is well above WHO's air quality guidelines for annual mean concentrations, with an interim target of 35 $\mu g/m^3$ and a final target of 10 $\mu g/m^3$.[14]

B. Regression Analyses

We first examine how well the number of people cooked for[15] (our main measure from the self-reported food diary) predicts time spent cooking, as measured by our iButtons (Table 2).[16] In the pooled regression (col. 5), each additional person cooked for predicts 0.54 hours of cooking per day

13 There were 376 measures of wood weights, but we dropped 17 negative values (4.5% of the data). The likely cause of these negative values is that the household added wood to the woodpile before it was weighed the following day.

14 See http://whqlibdoc.who.int/hq/2006/WHO_SDE_PHE_OEH_06.02_eng.pdf for more detail.

15 We use maximum number of people cooked for lunch or dinner.

16 Recall that our stove usage metrics generated from iButtons temperature data incorporate visual observations of stove use in the algorithm to convert temperatures to stove usage.

(95% confidence interval [CI] = -0.0043 to 1.1, $p < 0.1$). This point estimate is 6.4% of a standard deviation and about 5.0% of the mean of total hours cooked. Each additional lunch or dinner cooked predicts 3.3 more hours of stove use (95% CI = 0.30 to 6.4, $p < 0.05$). Finally, each time *matooke* or beans are cooked is associated with an increase in cooking time of 0.61 hours, but the 95% confidence interval is very wide (-0.77 to 2.0). While the coefficients are large in magnitude, the model's explanatory power is low ($R^2 = 0.073$).

When turning to within-household estimation (col. 6), we get useful variation only from homes that changed their cooking patterns substantially across the three measurement days. Thus, precision is typically much lower. Now the maximum number of people cooked for at lunch or dinner has a coefficient near zero (0.035 hours, 95% CI = -0.35 to 0.42). An extra main meal adds 1.3 hours of cooking (95% CI = -0.48 to 3.2). The coefficient on each time *matooke* or beans is cooked remains similar (0.66 hours, 95% CI = -0.028 to 1.4, $p < 0.1$), but now is marginally statistically significant. The Hausman test ($p = 0.46$) fails to reject the null hypothesis that the random effects estimator (or in our case the OLS with clustering estimator) is consistent and therefore is preferred over the fixed effect specification.

We next examine how well the number of people cooked for predicts kilograms of wood use (Table 3). In the pooled OLS regression (col. 5), adding one to the maximum number of people cooked for at lunch or dinner predicts an increase of wood use by 0.65 kilograms (CI = 0.39 to 0.90, $p < 0.01$). This point estimate is 11% of a standard deviation and about 6.8% of the mean daily wood use. Cooking an additional main meal (lunch or dinner) predicts a 2.0 kg increase in wood used (95% CI = 0.12 to 3.9, $p < 0.05$). The number of instances of cooking beans or *matooke* for lunch or dinner has no predictive power ($\beta = -0.28$, 95% CI = -0.95 to 0.39, $p = 0.41$). Again, while some coefficients are large, the explanatory power is low ($R^2 = 0.084$).

When we include a fixed effect for each household, the coefficients become smaller in magnitude and lose statistical significance. The Hausman test ($p = 0.015$) suggests that the fixed effect estimate is preferred.

We next examine how time spent cooking (as measured by iButtons) predicts kilograms of wood use (Table 4). In the pooled regression, 10 hours of additional cooking (about 1.2 times the standard deviation and about 90% of the mean) predicts 1.8 kilograms higher wood use (95% CI = 1.0 to 2.5; col. 1). This point estimate is about 31% of a standard deviation and 19% of the mean of wood use. The explanatory power ($R^2 = 0.10$) is consistent with measurement error in wood use, measurement error in time cooking, and with stoves varying substantially in wood consumption per hour cooking.

To account for possible measurement error in wood weights, we instrument hours of cooking with three instrumental variables from the food diary: the maximum number of people cooked for at lunch or dinner, number of main meals cooked per day (lunch and/or dinner), and number of instances of cooking beans or *matooke* per day. The two-stage least squares estimate implies a much higher coefficient on wood used per hour cooked. The estimate implies that on days a household cooked 10 additional hours, it used 5.7 kilograms more wood (95% CI = 2.5 to 8.9; col. 2). This point estimate is about three times larger than the OLS estimate in column 1.

When we include a fixed effect for each household, the estimate implies that on days a household cooks 10 additional hours, it uses 2.0 kilograms more wood (95% CI = -0.14 to 4.1; col. 3). The Hausman test does not reject that the OLS estimates are consistent. The instrumental variables approach in the fixed effects specification lacks a strong first stage, so we do not estimate that model.

We next separate out hours of cooking on what the family called its primary versus its secondary stove (col. 4). Hours on the main three-stone fire predict wood use ($\beta = 0.31$, 95% CI = 0.15 to 0.48), while hours cooked on the secondary three-stone fire have a much smaller value ($\beta = 0.01$, 95% CI = -0.16 to 0.18). We are not sure why the secondary stove point estimate is so close to zero. These results are perhaps consistent with a larger fire on the primary stove and the secondary stove being used for reheating a sauce or making a separate meal for a child or person on a restricted diet.

In the FE regression, we find no statistically significant correlations, although the point estimates on both stoves are close to the estimate of total hours (col. 5). There is neither a large nor a statistically significant effect of squared hours on either stove (col. 6 and 7).

We next examine how well the number of people cooked for predicts PM concentration, as measured by UCB-PATS (Table 5). Cooking for one additional person predicts a 9.3% increase in daily average PM concentration ($\beta = 0.089$, 95% CI = 0.031 to 0.15, $p < 0.01$; col. 5). Cooking an extra main meal (lunch or dinner) predicts a 23% higher daily average PM concentration, but the confidence interval is wide ($\beta = 0.20$, 95% CI = -0.16 to 0.56, $p = 0.27$; col. 5). The number of instances of cooking beans or *matooke* also had no statistically significant effect on daily average PM concentration. When we include a fixed effect for each household (col. 6), we find similar results. The Hausman test does not reject that the OLS is consistent.

We review how well time spent cooking (as measured by stove usage monitors) predicts PM concentration (as measured by UCB-PATS, Table 6). Pooling across homes, there is neither a large nor statistically significant effect of time spent cooking on average PM concentration ($\beta = 0.0083$, 95% CI = -0.012 to 0.028; col. 1). When we instrument for the maximum number of people cooked for at lunch or dinner, using the number of main meals cooked per day (lunch and/or dinner), and the number of instances of cooking beans or *matooke* per day, the estimate is much larger and statistically significant. An additional hour of cooking is associated with the 24-hour mean concentrations of PM increasing by 6.5% ($\beta = 0.063$, 95% CI = 0.0014 to 0.12; col. 2). This point estimate is over seven times larger than that in the pooled analysis.

When we include a fixed effect for each household, the estimate implies that an additional hour of cooking is associated with the 24-hour mean concentrations of PM increasing by 3.5% ($\beta = 0.034$, 95% CI = 0.0061 to 0.062; col. 3). The Hausman test does not reject that the OLS is consistent. The instrumental variables approach in the fixed effects specification lacks a strong first stage.

When examining how time spent cooking on the main stove (versus on the secondary stove) predicts PM concentrations, we find no statistically significant correlations with OLS (col. 4). When we include a fixed effect for each household, an additional hour of cooking on the secondary stove is associated with a 5.7% increase in mean 24-hour PM concentrations ($\beta = 0.055$, 95% CI = 0.0023 to 0.11; col. 5). However, the Hausman test does not reject that the OLS is consistent. There is no large or statistically significant effect of squared hours on either stove (col. 6 and 7).

Finally, we examine how a kilogram of wood used predicts PM concentrations, as measured by UCB-PATS (Table 7).[17] In the pooled regression, a 10% increase in wood used predicts a 3.0% increase in daily average PM concentration (β = 0.30, 95% CI = 0.093 to 0.50, col. 1). Although the effect is sizable, the explanatory power is low (R^2 = 0.039).

When instrumenting wood use with the maximum number of people cooked for at lunch or dinner, the number of main meals cooked per day (lunch and/or dinner), the number of instances of cooking beans or *matooke* per day, and hours main and secondary stove cooked adjusted by reliability, the estimate implies that a 10% increase in wood used predicts about a 5.7% increase in daily average PM concentration (β = 0.57, 95% CI = 0.017 to 1.1; col. 2). This point estimate is about twice that in the OLS analysis.

When we include a fixed effect for each household, the estimate implies that a 10% increase in wood used predicts a 1.6% increase in daily average PM concentration (β = 0.16, 95% CI = 0.022 to 0.30; col. 3). The Hausman test does not reject that the OLS is consistent. The instrumental variables approach in the fixed effects specification lacks a strong first stage.

V. Conclusions and Policy Implications

If we mismeasure stove usage, then carbon credits will be allocated incorrectly. In addition, our measures of how stoves affect outcomes such as health will be subject to unknown biases. These biases will also reduce our ability to understand what interventions might reduce harms from household air pollution. Our findings emphasize the importance of using multiple measures to understand cooking practices.

Of the main analyses, we find statistically significant positive correlations between five of the possible six pairs of proxies for cooking: estimated time spent cooking and number of people cooked for (weakly significant), kilograms of wood used and number of people cooked for (strongly significant), kilograms of wood used and estimated time spent cooking (strongly significant), PM concentrations and number of people cooked for (strongly significant), PM concentrations and estimated time spent cooking (not significant), and PM concentrations and kilograms of wood used (strongly significant). At the same time, the coefficients are often not very large and the explanatory power of each regression is low. Instrumental variable estimates adjust for measurement error and the coefficient estimates are typically several times larger than the OLS estimates.

We also included as covariates the number of main meals cooked, number of instances beans or *matooke* were cooked, hours main stove cooked (as well as centered and squared), and hours secondary stove cooked (as well as centered and squared). We found statistically significant correlations (all positive) between estimated time spent cooking and number of main meals cooked, kilograms of wood

[17] It is standard in the literature to analyze the natural log of PM. Thus, we use natural log of wood use to predict the natural log of PM so units are comparable. Additionally, results are similar analyzing kilograms of wood.

used and number of main meals cooked, and kilograms of wood used and hours main stove cooked. Within-household estimators eliminate the bias of time-invariant omitted variables (ventilation, altitude, etc.). However, when we controlled for these household fixed effects, estimates were imprecise because most households did not change their cooking very much from day to day.

Some variation in outcomes is due to our measures being conceptually distinct: wood use is not the same as PM concentration or hours of cooking. Additional variations in outcomes are due to variations in homes (e.g., ventilation), fuel (wet or dry), stoves (good or bad airflow), the weather (wind, temperature), and so forth. The very modest R^2 values we estimated are consistent with a substantial measurement error in most or all of our measures. We used instrumental variables to address the measurement error and get estimates that are more sensible in those cases. However, we cannot determine whether the measurement error was largely due to low reliability (random error) or low validity (bias).

These findings highlight challenges for researchers investigating health-related and other impacts of fuel-efficient cookstoves and for auditing carbon credits for fuel-efficient cookstoves. For example, when measures of fuel (and carbon) savings use only a single method for assessing stove use, such as a kitchen performance test, and use only a modest sample size, results can have substantial measurement errors. Given the weak correlations between measurements, we recommend that multiple measures should be used to increase the validity of estimated impacts.

In addition to the above findings, in a companion analysis we find that periods of intensive in-person monitoring of wood use significantly changes which cookstoves are used. These changes revert immediately when observers depart (Simons et al. 2017). Therefore, it is important to have low-cost, long-term and inconspicuous monitors to measure daily stove usage. Coupling SUMs with KPTs or other measurements may achieve this goal. For fuel-efficient cookstove programs to have the desired public effect of lowering carbon emissions and the desired private effect of improving household health, it is important to continue improving stove usage monitors and other stove measurement techniques.

Table 1: Daily hours cooked, particulate matter concentrations, and food diary data

Variable	Mean	Std. Dev.	Min.	Max.	N
Hours main stove cooked adjusted by reliability	6.44	6.04	0	21.48	219
Hours secondary stove cooked adj. by reliab.	4.51	5.19	0	20	219
Hours main and secondary stove cooked adj. by reliab.	10.95	8.54	0.14	36.55	219
Hours main stove cooked adj. by reliab., cntrd. and sq.	36.28	43.68	0.01	214.88	219
Hours secondary stove cooked adj. by reliab., cntrd. and sq.	27.05	41.66	0	264.85	219
Average Particulate Matter concentration ($\mu g/m^3$)	1018.77	1002.49	7.19	5548.01	365
Number of meals cooked per day	3.34	0.89	0	4	391
Number of main meals cooked per day (lunch and/or dinner)	1.85	0.4	0	2	400
Number of instances of beans or matooke per day	2.48	1.04	0	4	400
Net wood used daily (weight in kg)	9.91	6.66	0	46.72	359
Net wood used daily with top 5% coding (weight in kg)	9.62	5.72	0	24.27	359
Total number of people that breakfast was cooked for	5.37	3	0	14	403
Total number of people that lunch was cooked for	5.28	2.84	0	16	403
Total number of people that tea was cooked for	4.36	3.44	0	15	403
Total number of people that dinner was cooked for	6.09	2.38	0	12	403
Max number of people cooked for lunch or dinner	6.34	2.36	0	16	403

Notes: Data was collected on daily sequential visits to each household. In a few cases data was not retrieved until two days later, therefore we scale all measures to be proportional to use in a 24 hour period.

Hours cooked is derived from a predictive logistic regression of temperature data on observations of stoves in use. The process is described in detail in Simons *et al.* (2014). Reliability is the overall consistency of a measure.

Average Particulate Matter (PM) concentration is based on protocol for UCB Particle And Temperature Sensors produced by Berkeley Air Monitoring Group. WHO guideline PM concentration for $PM_{2.5}$ is an annual mean concentration of 35 $\mu g/m^3$ for the interim target 1.

Number of meals cooked per day is based on four potential meals: breakfast, lunch, tea, and dinner.

Net wood used is calculated after dropping 17 observations of negative wood weights, which likely occurred when households added wood to the designated pile before it was weighed.

Max number of people cooked for lunch or dinner takes the highest value of either lunch or dinner as those meals are the bulk of cooking.

Table 2: Estimated number of hours spent cooking (using iButtons) and food diaries
Dependent variable = No. of hours cooked daily adjusted by reliability

VARIABLES	(1) OLS	(2) FE	(3) OLS	(4) FE	(5) OLS	(6) FE
Max number of people cooked for lunch or dinner	0.657**	0.153	0.593**	0.0972	0.543*	0.0350
	(0.275)	(0.196)	(0.269)	(0.196)	(0.275)	(0.197)
Number of main meals cooked per day (lunch and/or dinner)			4.097***	2.012**	3.331**	1.341
			(1.201)	(0.857)	(1.525)	(0.918)
Number of instances of beans or matooke per day					0.609	0.664*
					(0.693)	(0.349)
Constant	6.876***		-0.256		-0.0344	
	(1.802)		(2.530)		(2.565)	
Observations	218	218	215	215	215	215
R-squared	0.034	0.005	0.069	0.047	0.073	0.075
Hausman Test (Prob > χ^2)		0.303		0.274		0.457
Number of household fixed effects		90		90		90

Robust standard errors adjusted for clustering by household in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: The Hausman test examines whether we can reject that the random effects estimator is consistent (in which case it will be more efficient than fixed effects), or whether we should only rely on the fixed effects estimator.
For definitions of variables see Table 1 footnotes.

Table 3: Daily wood used for cooking and food diaries
Dependent variable = kg. of wood used daily

VARIABLES	(1) OLS	(2) FE	(3) OLS	(4) FE	(5) OLS	(6) FE
Max number of people cooked for lunch or dinner	0.654***	-0.0752	0.626***	-0.112	0.645***	-0.107
	(0.136)	(0.247)	(0.133)	(0.251)	(0.130)	(0.253)
Number of main meals cooked per day (lunch and/or dinner)			1.620*	0.998	1.992**	1.061
			(0.882)	(0.916)	(0.948)	(1.022)
Number of instances of beans or matooke per day					-0.282	-0.0573
					(0.339)	(0.408)
Constant	5.471***		2.649		2.545	
	(0.906)		(1.837)		(1.810)	
Observations	358	358	357	357	357	357
R-squared	0.070	0.000	0.082	0.006	0.084	0.006
Hausman Test (Prob > χ^2)		0.004		0.007		0.015
Number of household fixed effects		152		152		152

Robust standard errors adjusted for clustering by household in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Note: The Hausman test examines whether we can reject that the random effects estimator is consistent (in which case it will be more efficient than fixed effects), or whether we should only rely on the fixed effects estimator.
For definitions of variables see Table 1 footnotes.

Table 4: Daily wood used for cooking and estimated number of hours spent cooking (using iButtons)
Dependent variable = kg. of wood used daily

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	OLS	IV	FE	OLS	FE	OLS	FE
Hours main stove cooked adjusted by reliability				0.314***	0.208	0.268***	0.0737
				(0.0833)	(0.173)	(0.0978)	(0.240)
Hours main stove cooked adj. by reliab., cntrd. and sq.						0.0108	0.0165
						(0.0121)	(0.0191)
Hours secondary stove cooked adj. by reliab.				0.00649	0.185	-0.0410	0.159
				(0.0854)	(0.211)	(0.141)	(0.302)
Hours secondary stove cooked adj. by reliab., cntrd. and sq.						0.00966	0.00677
						(0.0129)	(0.0278)
Hours main and secondary stove cooked adj. by reliab.	0.178***	0.574***	0.198*				
	(0.0376)	(0.163)	(0.107)				
Constant	6.468***			6.364***		6.225***	
	(0.504)			(0.504)		(0.517)	
Observations	196	194	196	196	196	196	196
R-squared	0.102		0.030	0.154	0.030	0.164	0.038
Hausman Test (Prob > χ^2)			0.867		0.656		0.880
Number of households fixed effects			85		85		85

Robust standard errors adjusted for clustering by household in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: The IV estimate in col. 2 used three instrumental variables: number of main meals cooked per day (lunch and/or dinner), number of instances of beans or matooke per day, and the maximum number of people cooked for lunch or dinner. The F statistic on the instruments in the first stage was 7.23 ($p < 0.01$; full results are in the appendix). The first stage for the fixed effects IV regression was weak ($F = 2.08$, $p = 0.11$), so we do not report fixed effect IV estimates.
The Hausman test examines whether we can reject that the random effects estimator is consistent (in which case it will be more efficient than fixed effects), or whether we should only rely on the fixed effects estimator.
For definitions of variables see Table 1 footnotes.

Table 5: Natural log of daily particulate matter (PM) concentrations and food diaries
Dependent variable = natural log of PM concentrations in micrograms per meter cubed

VARIABLES	(1) OLS	(2) FE	(3) OLS	(4) FE	(5) OLS	(6) FE
Max number of people cooked for lunch or dinner	0.0926***	0.0819***	0.0886***	0.0815***	0.0891***	0.0848***
	(0.0299)	(0.0259)	(0.0293)	(0.0264)	(0.0296)	(0.0268)
Number of main meals cooked per day (lunch and/or dinner)			0.194	0.0269	0.203	0.0641
			(0.160)	(0.0959)	(0.183)	(0.107)
Number of instances of beans or matooke per day					-0.00726	-0.0332
					(0.0637)	(0.0431)
Constant	5.916***		5.583***		5.580***	
	(0.201)		(0.354)		(0.356)	
Observations	364	364	361	361	361	361
R-squared	0.051	0.044	0.058	0.046	0.058	0.048
Hausman Test (Prob > χ^2)		0.446		0.068		0.138
Number of household fixed effects		148		148		148

Robust standard errors adjusted for clustering by household in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: The Hausman test examines whether we can reject that the random effects estimator is consistent (in which case it will be more efficient than fixed effects), or whether we should only rely on the fixed effects estimator.
For definitions of variables see Table 1 footnotes.

Table 6: Natural log of daily PM concentrations and estimated number of hours spent cooking (using iButtons)
Dependent variable = natural log of PM concentrations in micrograms per meter cubed

VARIABLES	(1) OLS	(2) IV	(3) FE	(4) OLS	(5) FE	(6) OLS	(7) FE
Hours main stove cooked adjusted by reliability				-0.00455	0.0156	-0.00802	0.0327
				(0.0139)	(0.0243)	(0.0186)	(0.0277)
Hours main stove cooked adj. by reliab., cntrd. and sq.						0.00110	-0.00322
						(0.00200)	(0.00259)
Hours secondary stove cooked adj. by reliab.				0.0239	0.0553**	0.0437	0.0766**
				(0.0201)	(0.0268)	(0.0291)	(0.0316)
Hours secondary stove cooked adj. by reliab., cntrd. and sq.						-0.00315	-0.00449
						(0.00361)	(0.00349)
Hours main and secondary stove cooked adj. by reliab.	0.00826	0.0630**	0.0341**				
	(0.00994)	(0.0314)	(0.0141)				
Constant	6.320***			6.326***		6.304***	
	(0.143)			(0.143)		(0.142)	
Observations	203	199	203	203	203	203	203
R-squared	0.006		0.047	0.018	0.054	0.027	0.082
Hausman Test (Prob > χ^2)			0.127		0.312		0.443
Number of Households			84		84		84

Robust standard errors adjusted for clustering by household in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: The IV estimate in col. 2 used three instrumental variables: number of main meals cooked per day (lunch and/or dinner), number of instances of beans or matooke per day, and the maximum number of people cooked for lunch or dinner. The F statistic on the instruments in the first stage was 6.22 ($p < 0.01$; full results are in the appendix). The first stage for the fixed effects IV regression was weak ($F = 1.80$, $p = 0.15$), so we do not report fixed effect IV estimates.
The Hausman test examines whether we can reject that the random effects estimator is consistent (in which case it will be more efficient than fixed effects), or whether we should only rely on the fixed effects estimator.
For definitions of variables see Table 1 footnotes.

Table 7: Natural log of daily PM concentrations and natural log of wood used for cooking

Dependent variable = natural log of PM concentrations in micrograms per meter cubed

VARIABLES	(1) OLS	(2) IV	(3) FE
Log of wood used daily with top 5% coding (weight in kg)	0.296***	0.569**	0.159**
	(0.102)	(0.282)	(0.0694)
Constant	5.893***		
	(0.222)		
Observations	325	179	325
R-squared	0.039		0.028
Hausman Test (Prob > χ^2)			0.561
Number of household fixed effects			141

Robust standard errors adjusted for clustering by household in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: The IV estimate in col. 2 used four instrumental variables: number of main meals cooked per day (lunch and/or dinner), number of instances of beans or matooke per day, the maximum number of people cooked for lunch or dinner, and hours main and secondary stove cooked adj. by reliab.. The F statistic on the instruments in the first stage was 11.48 ($p < 0.01$; full results are in the appendix). The first stage for the fixed effects IV regression was weak ($F = 1.11$, $p = 0.36$), so we do not report fixed effect IV estimates.

The Hausman test examines whether we can reject that the random effects estimator is consistent (in which case it will be more efficient than fixed effects), or whether we should only rely on the fixed effects estimator.

For definitions of variables see Table 1 footnotes.

Appendix 1: First stages of IV regressions
Dependent variable = (1) and (2) No. of hours cooked daily adjusted by reliability and (3) Log of wood used daily with top 5% coding

	(1)	(2)	(3)
VARIABLES	IV first stage	IV first stage	IV first stage
Number of main meals cooked per day (lunch and/or dinner)	3.927**	3.657**	0.450***
	(1.633)	(1.517)	(0.136)
Number of instances of beans or matooke per day	0.674	0.569	-0.0683
	(0.720)	(0.696)	(0.0460)
Max number of people cooked for lunch or dinner	0.655**	0.512*	0.0706***
	(0.303)	(0.285)	(0.0189)
Hours main and secondary stove cooked adj. by reliab.			0.0137***
			(0.00500)
Constant	-1.764	-0.826	0.699***
	(2.808)	(2.626)	(0.259)
Observations	194	199	179
R-squared	0.094	0.079	0.206

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Note: Columns 1, 2, and 3 refer respectively to tables 6, 8, and 9.

Bibliography

Akaike, Hirotugu. 1981. "Likelihood of a Model and Information Criteria." *Journal of Econometrics* 16: 3–14.

Alexander, Donee A., Amanda Northcross, Theodore Karrison, Oludare Morhasson-Bello, Nathaniel Wilson, Omolola M. Atalabi, Anindita Dutta, et al. 2018. "Pregnancy Outcomes and Ethanol Cook Stove Intervention: A Randomized-Controlled Trial in Ibadan, Nigeria." *Environment International* 111: 152–163. https://doi.org/10.1016/j.envint.2017.11.021.

Anenberg, Susan C., Kalpana Balakrishnan, James Jetter, Omar Masera, Sumi Mehta, Jacob Moss, and Veerabhadran Ramanathan. 2013. "Cleaner cooking solutions to achieve health, climate, and economic cobenefits." *Environ. Sci. Technol.*: 3944-3952.

Arnold, J.E. Michael, Gunnar Köhlin, and Reidar Persson. 2006. "Woodfuels, Livelihoods, and Policy Interventions: Changing Perspectives." *World Development* 34 (3): 596–611. doi:10.1016/j.worlddev.2005.08.008.

Ashraf, Nava, James Berry, and Jesse M Shapiro. 2010. "Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia." *American Economic Review* 100 (5): 2383–2413. https://doi.org/10.1257/aer.100.5.2383.

Bailis, Robert, Kirk R. Smith, and Rufus Edwards. 2007. "Kitchen Performance Test (KPT)." Univerisity of California, Berkeley, CA.

Bailis, Robert, Majid Ezzati, and Daniel M Kammen. 2005. "Mortality and Greenhouse Gas Impacts of Biomass and Petroleum Energy Futures in Africa." *Science* 308 (5718): 98–103. doi:10.1126/science.1106881.

Bailis, Robert, Rudi Drigo, Adrian Ghilardi, and Omar Masera. 2015. "The Carbon Footprint of Traditional Woodfuels." *Nature Climate Change* 5 (3): 266–272. https://doi.org/10.1038/nclimate2491.

Beltramo, Theresa, Garrick Blalock, David I. Levine, and Andrew M. Simons. 2015a. "Does Peer Use Influence Adoption of Efficient Cookstoves? Evidence From a Randomized Controlled Trial in Uganda." *Journal of Health Communication* 20 (S1): 55–66. https://doi.org/10.1080/10810730.2014.994244.

Beltramo, Theresa, Garrick Blalock, David I. Levine, and Andrew M. Simons. 2015b. "The Effect of Marketing Messages and Payment over Time on Willingness to Pay for Fuel-Efficient Cookstoves." *Journal of Economic Behavior & Organization* 118: 333–45. https://doi.org/10.1016/j.jebo.2015.04.025.

Bensch, Gunther, and Jörg Peters. 2015. "The Intensive Margin of Technology Adoption – Experimental Evidence on Improved Cooking Stoves in Rural Senegal." *Journal of Health Economics* 42: 44–63. https://doi.org/10.1016/j.jhealeco.2015.03.006.

Berkeley Air Monitoring Group, and Berkeley Indoor Air Pollution Team, School of Public Health, University of California. 2010. "UCB Particle Monitor User Manual." Berkeley, CA.

http://edge.rit.edu/edge/P13625/public/Reference%20Documents/User%20Manual_UCB%20Particle%20Monitor_v8%2030Dec2010.pdf.

Bernheim, B. Douglas, Jonathan Meer, and Neva K. Novarro. 2016. "Do consumers exploit commitment opportunities? Evidence from natural experiments involving liquor consumption." *American Economic Journal: Economic Policy* 8, no. 4: 41-69.

Bond, Tami, Chandra Venkataraman, and Omar Masera. 2004. "Global Atmospheric Impacts of Residential Fuels." *Energy for Sustainable Development* 8 (3): 20–32.

Bonjour, Sophie, Heather Adair-Rohani, Jennyfer Wolf, Nigel G. Bruce, Sumi Mehta, Annette Prüss-Ustün, Maureen Lahiff, Eva A. Rehfuess, Vinod Mishra, and Kirk R. Smith. 2013. "Solid Fuel Use for Household Cooking: Country and Regional Estimates for 1980-2010." *Environmental Health Perspectives* 121 (7): 784–790. https://doi.org/10.1289/ehp.1205987.

Bryan, Gharad, Dean Karlan, and Scott Nelson. 2010. "Commitment devices." *Annu. Rev. Econ.* 2, no. 1: 671-698.

Burke, Paul J, and Guy Dundas. 2015. "Female Labor Force Participation and Household Dependence on Biomass Energy: Evidence from National Longitudinal Data." *World Development* 67: 424–37. doi:10.1016/j.worlddev.2014.10.034.

Chay, Kenneth Y., and Michael Greenstone. 2003. "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): 1121–1167. https://doi.org/10.1162/00335530360698513.

Chowdhury, Zohir, Rufus Edwards, Michael Johnson, Kyra Naumoff Shields, Tracy Allen, Eduardo Canuz, and Kirk R Smith. 2007. "An Inexpensive Light-Scattering Particle Monitor: Field Validation." *Journal of Environmental Monitoring* 9 (10): 1099–1106.

Cohen, Jessica, and Pascaline Dupas. 2010. "Free Distribution or Cost-Sharing? Evidence from a Malaria Prevention Experiment." *Quarterly Journal of Economics* 125 (1): 1–45.

Currie, J, and WR Walker. 2009. "Traffic Congestion and Infant Health: Evidence from E-Zpass." *American Economic Journal: Applied Economics* 3 (1): 65–90. http://www.nber.org/papers/w15413.

Dickinson, Katherine L, Ernest Kanyomse, Ricardo Piedrahita, Evan Coffey, Isaac J Rivera, James Adoctor, Rex Alirigia, et al. 2015. "Research on Emissions, Air Quality, Climate, and Cooking Technologies in Northern Ghana (Reaccting): Study Rationale and Protocol." *BMC Public Health* 15 (1): 126. https://doi.org/10.1186/s12889-015-1414-1.

Dupas, Pascaline. 2014. "Getting Essential Health Products to Their End Users: Subsidize, but How Much?" *Science* 345 (6202): 1279–81. https://doi.org/10.1126/science.1256973.

Ezzati, Majid, and Daniel M Kammen. 2001. "Indoor Air Pollution from Biomass Combustion and Acute Respiratory Infections in Kenya: An Exposure-Response Study." *Lancet* 358 (9282): 619–24. http://www.ncbi.nlm.nih.gov/pubmed/11530148.

Ezzati, Majid, and Daniel M Kammen. 2002. "Evaluating the Health Benefits of Transitions in Household Energy Technologies in Kenya." *Energy Policy* 30 (10): 815–26. https://doi.org/10.1016/S0301-4215(01)00125-2.

Ezzati, Majid, Homayoun Saleh, and Daniel M Kammen. 2000. "The Contributions of Emissions and Spatial Microenvironments to Exposure to Indoor Air Pollution from Biomass Combustion in Kenya." *Environmental Health Perspectives* 108 (9): 833–39.

Fischer, Greg, Dean Karlan, Margaret McConnell, and Pia Raffler. 2019. "Short-Term Subsidies and Seller Type: A Health Products Experiment in Uganda." *Journal of Development Economics* 137: 110–24. https://doi.org/doi.org/10.1016/j.jdeveco.2018.07.013.

Garland, C., C. F. Gould, and D. Pennise. 2018. "Usage and Impacts of the Envirofit HM-5000 Cookstove." *Indoor Air* 28 (4): 640–650. https://doi.org/10.1111/ina.12460.

GBD 2017, "GBD Results Tools", http://ghdx.healthdata.org/gbd-results-tool, accessed on 23rd March 2019.

Gould, Carlos F., and Johannes Urpelainen. 2018. "LPG as a clean cooking fuel: Adoption, use, and impact in rural India." *Energy Policy* 122: 395-408.

Graham, Eric a., Omkar Patange, Martin Lukac, Lokendra Singh, Abhishek Kar, Ibrahim H. Rehman, and Nithya Ramanathan. 2014. "Laboratory Demonstration and Field Verification of a Wireless Cookstove Sensing System (WiCS) for Determining Cooking Duration and Fuel Consumption." *Energy for Sustainable Development* 23: 59–67. doi:10.1016/j.esd.2014.08.001.

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

Harrell, Stephen, Theresa Beltramo, Garrick Blalock, Juliet Kyayesimira, David I. Levine, and Andrew M. Simons. 2016. "What Is a Meal?: Comparing Methods of Auditing Carbon Offset Compliance for Fuel Efficient Cookstoves." *Ecological Economics* 128: 8–16.

Himmler, Oliver, Robert Jäckle, and Philipp Weinschenk. 2019. "Soft Commitments, Reminders, and Academic Performance." *American Economic Journal: Applied Economics* 11, no. 2: 114-42.

Jain, Abhishek, S. Ray, K. Ganesan, M. Aklin, C. Cheng, and J. Urpelainen. 2018. "Access to clean cooking energy and electricity." *Council on Energy, Environment and Water (CEEW), India:* https://www.ceew.in/sites/default/files/CEEW-Access-to-Clean-Cooking-Energy-and-Electricity-11Jan19_0.pdf.

Johnson MA, Chiang RA. 2015. Quantitative Guidance for Stove Usage and Performance to Achieve Health and Environmental Targets. *Environ Health Perspect*. doi: 10.1289/ehp.1408681.

Kammen, Daniel M, Robert Bailis, and Antonia Herzog. 2002. "Clean Energy for Development and Economic Growth: Biomass and Other Renewable Energy Options to Meet Energy and Development Needs in Poor Nations." New York: UNDP and Government of Morocco.

Karlan, Dean, and Leigh L. Linden. 2018. Loose knots: Strong versus weak commitments to save for education in Uganda. No. w19863. *National Bureau of Economic Research*.

Kast, Felipe, Stephan Meier, and Dina Pomeranz. 2018. "Saving more in groups: Field experimental evidence from Chile." *Journal of Development Economics* 133: 275-294.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. 2015. "Self-control at work." *Journal of Political Economy* 123, no. 6: 1227-1277.

Krall, E A, and J T Dwyer. 1987. "Validity of a Food Frequency Questionnaire and a Food Diary in a Short-Term Recall Situation." *Journal of the American Dietetic Association* 87 (10): 1374–77.

Kremer, Michael, Gautam Rao, and Frank Schilbach. 2019. "Behavioral development economics." *Handbook of Behavioral Economics* 2.

LaFave, Daniel, Abebe Damte Beyene, Randall Bluffstone, Sahan T.M. Dissanayake, Zenebe Gebreegziabher, Alemu Mekonnen, and Michael Toman. 2019. "Impacts of Improved Biomass Cookstoves on Child and Adult Health: Experimental Evidence from Rural Ethiopia." Waterville, ME. http://web.colby.edu/drlafave/lafaveetal_stovehealth/.

Levine, David I, Theresa Beltramo, Garrick Blalock, Carolyn Cotterman, and Andrew M. Simons. 2018. "What Impedes Efficient Adoption of Products? Evidence from Randomized Sales Offers for Fuel-Efficient Cookstoves in Uganda." *Journal of the European Economic Association* 16 (6): 1850–80.

Lee, Carrie M, Chelsea Chandler, Michael Lazarus, and Francis X Johnson. 2013. "Assessing the Climate Impacts of Cookstove Projects: Issues in Emissions Accounting." Stockholm.

Levine, David I., Theresa Beltramo, Garrick Blalock, Carolyn Cotterman, and Andrew M. Simons. 2018. "What impedes efficient adoption of products? Evidence from randomized sales offers for fuel-efficient cookstoves in Uganda." *Journal of the European Economic Association* 16, no. 6: 1850-1880.

Lewis, Jessica J, and Subhrendu K Pattanayak. 2012. "Who Adopts Improved Fuels and Cookstoves? A Systematic Review." *Environmental Health Perspectives* 120 (5): 637–45. https://doi.org/10.1289/ehp.1104194.

Lim, Stephen S, Theo Vos, Abraham D Flaxman, Goodarz Danaei, Kenji Shibuya, Heather Adair-Rohani, Markus Amann, et al. 2012. "A Comparative Risk Assessment of Burden of Disease and Injury Attributable to 67 Risk Factors and Risk Factor Clusters in 21 Regions, 1990-2010: A Systematic Analysis for the Global Burden of Disease Study 2010." *Lancet* 380 (9859): 2224–60. http://www.ncbi.nlm.nih.gov/pubmed/23245609.

Masera, Omar R, Barbara D Saatkamp, and Daniel M Kammen. 2000. "From Linear Fuel Switching to Multiple Cooking Strategies: A Critique and Alternative to the Energy Ladder Model." *World Development* 28 (12): 2083–2103. https://doi.org/10.1016/S0305-750X(00)00076-0.

McCracken, John, J Schwartz, M Mittleman, L Ryan, A Diaz Artiga, and Kirk R Smith. 2007. "Biomass Smoke Exposure and Acute Lower Respiratory Infections Among Guatemalan Children." *Epidemiology* 18 (5): S185–S185.

Mikhed, Vyacheslav, Barry Scholnick, and Hyungsuk Byun. 2017. "Spatial Commitment Devices and Addictive Goods: Evidence from the Removal of Slot Machines from Bars." *FRB of Philadelphia Working Paper No. 17-34*. Available at SSRN: https://ssrn.com/abstract=3049206.

Miller, Grant, and A. Mushfiq Mobarak. 2013. "Gender Differences in Preferences, Intra-Household Externalities, and Low Demand for Improved Cookstoves." Cambridge, MA.

Mortimer, Kevin, Chifundo B Ndamala, Andrew W Naunje, Jullita Malava, Cynthia Katundu, William Weston, Deborah Havens, et al. 2017. "A Cleaner Burning Biomass-Fuelled Cookstove Intervention to Prevent Pneumonia in Children under 5 Years Old in Rural Malawi (the Cooking and Pneumonia Study): A Cluster Randomised Controlled Trial." *The Lancet* 389 (10065): 167–75. https://doi.org/10.1016/S0140-6736(16)32507-7.

Mukhopadhyay, Rupak, Sankar Sambandam, Ajay Pillarisetti, Darby Jack, Krishnendu Mukhopadhyay, Kalpana Balakrishnan, Mayur Vaswani, et al. 2012. "Cooking Practices, Air Quality, and the Acceptability of Advanced Cookstoves in Haryana, India: An Exploratory Study to Inform Large-Scale Interventions." *Global Health Action* 5: 19016–19016.

Piedrahita, Ricardo, Katherine L. Dickinson, Ernest Kanyomse, Evan Coffey, Rex Alirigia, Yolanda Hagar, Isaac Rivera et al. 2016. "Assessment of cookstove stacking in Northern Ghana using surveys and stove use monitors." *Energy for Sustainable Development* 34: 67-76.

Pillarisetti, Ajay, Makarand Ghorpade, Sathish Madhav, Arun Dhongade, Sudipto Roy, Kalpana Balakrishnan, Sambandam Sankar et al. 2019. "Promoting LPG usage during pregnancy: A pilot study in rural Maharashtra, India." *Environment international* 127: 540-549.

Pillarisetti, Ajay, Mayur Vaswani, Darby Jack, Kalpana Balakrishnan, Michael N Bates, Narendra K Arora, and Kirk R Smith. 2014. "Patterns of Stove Usage after Introduction of an Advanced Cookstove: The Long-Term Application of Household Sensors." *Environmental Science & Technology* 48 (24): 14525–33.

Pope III, C. Arden, Richard T. Burnett, Michael J. Thun, Eugenia E. Calle, Daniel Krewski, and George D. Thurston. 2002. "Lung Cancer, Cardiopulmonary Mortality, and Long-Term Exposure to Fine Particulate Air Pollution." *Journal of the American Medical Association* 287 (9): 1132–1141. https://doi.org/10.1001/jama.287.9.1132.

Prentice, Ross L. 2003. "Dietary Assessment and the Reliability of Nutritional Epidemiology Reports." *Lancet* 362 (9379): 182–83.

Ramanathan, Veerabhadran, and Gregory Carmichael. 2008. "Global and Regional Climate Changes due to Black Carbon." *Nature Geoscience* 1 (4): 221–27.

Rehfuess, Eva A., David J. Briggs, Mike Joffe, and Nicky Best. 2010. "Bayesian Modelling of Household Solid Fuel Use: Insights towards Designing Effective Interventions to Promote Fuel Switching in Africa." *Environmental Research* 110 (7): 725–732. https://doi.org/10.1016/j.envres.2010.07.006.

Royer, Heather, Mark Stehr, and Justin Sydnor. 2015. "Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company." *American Economic Journal: Applied Economics* 7, no. 3: 51-84.

Ruiz-Mercado, Ilse, Eduardo Canuz, and Kirk R. Smith. 2012. "Temperature Dataloggers as Stove Use Monitors (SUMs): Field Methods and Signal Analysis." *Biomass and Bioenergy* 47: 459–68. doi:10.1016/j.biombioe.2012.09.003.

Ruiz-Mercado, Ilse, Eduardo Canuz, Joan L. Walker, and Kirk R. Smith. 2013. "Quantitative Metrics of Stove Adoption Using Stove Use Monitors (SUMs)." *Biomass and Bioenergy* 57 (October): 136–48.

Ruiz-Mercado, Ilse, Nick L Lam, Eduardo Canuz, Gilberto Davila, and Kirk R Smith. 2008. "Low-Cost Temperature Loggers as Stove Use Monitors (SUMs)." *Boiling Point* 55: 16–18.

Ruiz-Mercado, Ilse, Omar Masera, Hilda Zamora, and Kirk R. Smith. 2011. "Adoption and Sustained Use of Improved Cookstoves." *Energy Policy* 39 (12). Elsevier: 7557–66. doi:10.1016/j.enpol.2011.03.028.

Simon, Gregory L, Adam G Bumpus, and Philip Mann. 2012. "Win-Win Scenarios at the Climate–development Interface: Challenges and Opportunities for Stove Replacement Programs through Carbon Finance." *Global Environmental Change* 22 (1): 275–87. doi:10.1016/j.gloenvcha.2011.08.007.

Simons, Andrew M., Theresa Beltramo, Garrick Blalock, and David I. Levine. 2014. "Comparing Methods for Signal Analysis of Temperature Readings from Stove Use Monitors." *Biomass and Bioenergy* 70: 476–88.

Simons, Andrew M., Theresa Beltramo, Garrick Blalock, and David I. Levine. 2017. "Using unobtrusive sensors to measure and minimize Hawthorne effects: Evidence from cookstoves." *Journal of Environmental Economics and Management* 86: 68-80.

Simons, Andrew M., Theresa Beltramo, Garrick Blalock, and David I. Levine. 2018. "Sensor Data to Measure Hawthorne Effects in Cookstove Evaluation." *Data in Brief* 18: 1334–1339. https://doi.org/10.1016/j.dib.2018.04.021.

Smith, Kirk R, and Ajay Pillarisetti. 2017. "Household Air Pollution from Solid Cookfuels and Its Effects on Health." *Injury Prevention and Environmental Health*: 133.

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. doi:10.1038/jes.2009.30.

Smith, Kirk R, John P McCracken, Martin W Weber, Alan Hubbard, Alisa Jenny, Lisa M Thompson, John Balmes, Anaité Diaz, Byron Arana, and Nigel Bruce. 2011. "Effect of Reduction in Household Air Pollution on Childhood Pneumonia in Guatemala (RESPIRE): A Randomised Controlled Trial." *Lancet* 378 (9804): 1717–26. https://doi.org/10.1016/S0140-6736(11)60921-5.

Smith, Kirk R, Nigel Bruce, Byron Arana, Alisa Jenny, Asheena Khalakdina, John McCracken, Anaite Diaz, Morten Schei, and Sandy Gove. 2006. "Conducting The First Randomized Control Trial of Acute Lower Respiratory Infections and Indoor Air Pollution: Description of Process and Methods." *Epidemiology* 17 (6): S44–S44.

Smith-Sivertsen, Tone, Esperanza Díaz, Dan Pope, Rolv T Lie, Anaite Díaz, John McCracken, Per Bakke, Byron Arana, Kirk R Smith, and Nigel Bruce. 2009. "Effect of Reducing Indoor Air Pollution on Women's Respiratory Symptoms and Lung Function: The RESPIRE Randomized Trial, Guatemala." *American Journal of Epidemiology* 170 (2): 211–20. doi:10.1093/aje/kwp100.

Sorrell, Steve, John Dimitropoulos, and Matt Sommerville. 2009. "Empirical Estimates of the Direct Rebound Effect: A Review." *Energy Policy* 37 (4): 1356–71. https://doi.org/10.1016/j.enpol.2008.11.026.

Thaler, Richard H., and Shlomo Benartzi. 2004. "Save more tomorrow™: Using behavioral economics to increase employee saving." *Journal of Political Economy* 112, no. S1: S164-S187.

The Gold Standard Foundation. 2013. "The Gold Standard: Simplified Methodology for Efficient Cookstoves." Geneva-Cointrin, Switzerland.

Thomas, Evan A, Christina K Barstow, Ghislaine Rosa, Fiona Majorin, and Thomas Clasen. 2013. "Use of Remotely Reporting Electronic Sensors for Assessing Use of Water Filters and Cookstoves in Rwanda." *Environmental Science & Technology* 47 (23): 13602–10. doi:10.1021/es403412x.

Thomas, Evan A, Sarita Tellez-Sanchez, Carson Wick, Miles Kirby, Laura Zambrano, Ghislaine Abadie Rosa, Thomas F. Clasen, and Corey Nagel. 2016. "Behavioral Reactivity Associated With Electronic Monitoring of Environmental Health Interventions: A Cluster Randomized Trial with Water Filters and Cookstoves." *Environmental science & technology* 50, no. 7: 3773-3780.

Tripathi, Alok and Ambuj Sagar. 2019. "Ujjwala, V2.0 What should be done next?" *Collaborative Clean Air Policy Centre*: https://ccapc.org.in/policy-briefs/2019/6/15/ujjwala-v2-what-should-be-done-next.

World Health Organization. 2006. "WHO Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide: Global Update 2005: Summary of Risk Assessment," 1–22. https://doi.org/10.1016/0004-6981(88)90109-6.