WP 2016-08

May 2016

Working Paper

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

Using Unobtrusive Sensors to Measure and Minimize Hawthorne Effects: Evidence from Cookstoves

**Andrew M. Simons, Theresa Beltramo,
Garrick Blalock, David I. Levine**

Using Unobtrusive Sensors to Measure and Minimize Hawthorne Effects:

Evidence from Cookstoves

Andrew M. Simons, Theresa Beltramo, Garrick Blalock, David I. Levine*

People act differently when they know they are being observed. This phenomenon, the Hawthorne effect, can bias estimates of program impacts. Unobtrusive sensors substituting for human observation can alleviate this bias. To demonstrate this potential, we used temperature loggers to measure fuel-efficient cookstoves as a replacement for three-stone fires. We find a large Hawthorne effect: when in-person measurement begins, participants increase fuel-efficient stove use approximately three hours/day (54%) and reduce three-stone fire use by approximately two hours/day (30%). When in-person measurement ends, participants reverse those changes. We then examine how this Hawthorne effect biases estimates of fuel use and particulate matter concentrations. Our results reinforce concerns about Hawthorne effects, especially in policy-relevant impact evaluations. We demonstrate that sensors can sometimes provide a solution.

Keywords: observation bias, Hawthorne effect, sensors, improved cookstoves, monitoring and evaluation, impact evaluation

* Simons and Blalock: Charles H. Dyson School of Applied Economics and Management, Cornell University (email: ams727@cornell.edu and garrick.blalock@cornell.edu). Beltramo: United Nations High Commissioner for Refugees, Geneva (email: beltramo@unhcr.org). Levine: Haas School of Business, University of California, Berkeley (email: levine@haas.berkeley.edu).

Acknowledgments: This study was funded by the United States Agency for International Development under Translating Research into Action, Cooperative Agreement No. GHS-A-00-09-00015-00. The recipient of the grant was Impact Carbon who co-funded and managed the project. Juliet Kyaesimira and Stephen Harrell expertly oversaw field operations and Amy Gu provided excellent research support. We thank Impact Carbon partners Matt Evans, Evan Haigler, Jimmy Tran, Caitlyn Toombs, and Johanna Young; U.C. Berkeley Household Energy, Climate, and Health Research Group partners including Kirk Smith, Ilse Ruiz-Mercado, and Ajay Pillarisetti; Berkeley Air partners including Dana Charron, David Pennise, Michael Johnson, and Erin Milner; the USAID TRAction Technical Advisory Group, and seminar participants at Cornell University, UCLA, and Oxford University for valuable comments. Data collection was carried out by the Center for Integrated Research and Community Development (CIRCODU), and the project's success relied on expert oversight by CIRCODU's Director General Joseph Ndemere Arineitwe and field supervisors Moreen Akankunda, Innocent Byaruhanga, Fred Isabirye, Noah Kirabo, and Michael Mukembo. We thank the Atkinson Center for a Sustainable Future at Cornell University, the Institute for the Social Sciences at Cornell University and the Cornell Population Center for additional funding of related expenses. The findings of this study are the sole responsibility of the authors, and do not necessarily reflect the views of their respective institutions, nor USAID or the United States Government.

Introduction

The validity of empirical research depends on data quality. Unlike the physical sciences, for which data often is generated in controlled laboratory settings, the social sciences typically measure variables involving human behaviors, which make data quality a challenge. Respondents often do not answer surveys candidly (Bertrand and Mullainathan 2001) and the act of surveying can change later behaviors of those being surveyed (Zwane et al. 2011). These drawbacks to surveys have been one factor contributing to a push for more experiments in environmental economics (Greenstone and Gayer 2009) and social science research more generally (Falk and Heckman 2009; Banerjee and Duflo 2009; Duflo, Glennerster, and Kremer 2008). While much has been learned from experiments in environmental economics, these types of experiments measuring human behaviors are susceptible to issues such as observation bias, or Hawthorne effect.

We explore an emerging class of technology—small, inexpensive, and unobtrusive sensors—as a remedy to the Hawthorne effect. A growing variety of sensors have become available to researchers. GPS trackers and motion detectors, for example, allow non-obtrusive measurement of subject location and body movements (Ermes et al. 2008). Medical doctors wear sensors that detect the scent of alcohol used in hand sanitizers to alert the doctor and/or patient if the doctor has not washed his or her hands recently (E. Smith 2014; Srigley et al. 2014). Loop detectors installed in the lanes of freeways allow monitoring of congestion and driver behavior (Bento et al. 2014).

The degree to which these sensors interfere with subjects' behavior can vary widely. In some cases, individuals may choose to be observed to motivate their own behavioral response. For example, long-distance bikers and runners can opt into programs that will report the location, time, and speed of excursions to a website that others can monitor (Mueller et al. 2010). Users of such schemes typically hope peer observation will increase their motivation. In other cases, such as room occupancy detectors that control lighting and climate control, the sensor may be far harder to notice (Buchanan, Russo, and Anderson 2014).

A major challenge for direct observational studies is that they alter participants' behavior. The effects of observers have been noted in cookstove studies (Ezzati, Saleh, and Kammen 2000; Smith-Sivertsen et al. 2009), in energy consumption (Schwartz et al. 2013), in public health (Clasen et al. 2012; Das, Hammer, and Leonard 2008; Leonard and Masatu 2006; Srigley et al. 2014) in development economics (Leonard 2008; Leonard and Masatu 2010; Muralidharan and Sundararaman 2010) and in social sciences more broadly (Levitt and List 2007; Levitt and List 2011). We demonstrate a technique to remedy the Hawthorne effect that uses unobtrusive temperature sensors in an evaluation of fuel-efficient cookstoves in Uganda. We use minimally invasive temperature sensors to measure usage of the fuel-efficient cookstoves and of the traditional three-stone fires.[1] We then compare usage of each stove in periods when observers visit the households with periods when no observers are present. We find a large Hawthorne effect:

[1] A three-stone fire is simply three large stones, approximately the same height, on which a cooking pot is balanced over a fire.

households increase the use of the fuel-efficient stove and decrease the use of three-stone fires on days they expect observers.

The observers visited homes to measure wood use and household exposure to particulate matter. Unfortunately, changes in cooking practices due to the observers will bias measures of wood use and of exposure to particulates. Fortunately, once the magnitude of this Hawthorne effect is known, we can estimate unbiased impacts of how fuel-efficient stoves affect wood use and exposure to particulate matter.

Data and methodology

We implemented a series of studies in rural areas of the Mbarara District in southwestern Uganda from February to September 2012, which focused on the adoption, and use of fuel-efficient stoves. At baseline almost all families cooked on a traditional three-stone fire (97%), usually located within a separate enclosed cooking hut. We introduced an Envirofit G-3300 stove. Its manufacturer reports that it uses 50% less fuel and reduces smoke and harmful gasses by 51% compared to a three stone fire (Envirofit Inc. 2011). The study area is characterized by agrarian livelihoods including raising livestock and farming *matooke* (starchy cooking banana), potatoes, and millet.

We tracked stove usage before and after the purchase of a fuel-efficient stove at 168 households spread across fourteen rural parishes in Mbarara.[2] Upon arriving in a new parish, staff displayed the fuel-efficient stove (Envirofit G-3300) and offered it

[2] The population of most Ugandan rural parishes ranges from 4,000 to 6,000.

for sale to anyone who wanted to purchase at 40,000 Ugandan Shillings (approximately USD $16, see Beltramo et al. (2015) and Levine et al. (2016) for an overview of the sales contract). Households were eligible to participate in the impact study if they mainly used wood as a fuel source, regularly cooked for eight or fewer persons, someone was generally home every day, and cooking was largely in an enclosed kitchen.

Eligible households who wanted to buy the stove were randomly assigned to two groups: early buyers, late buyers. Because it is crucial to measure both the use of the new stove *and* any reduction in use of traditional stoves (Ruiz-Mercado et al. 2011), we asked both early buyers and late buyers if they would agree to have stove use monitors (SUMs) that read stove temperatures placed on their traditional and Envirofit stoves. After giving consent, three stone fires were fitted with SUMs immediately and we collected a baseline round of data with only three stone fires present in homes.

Approximately two to three weeks later the early buyers group received their first Envirofit stove. We did a midline round of data collection that is not used in this study (but will be the basis of an impact evaluation, when the data are cleaned). Approximately five to six weeks later the late buyers received their first Envirofit stove. About six weeks after late buyers received their Envirofits, both groups were surprised with a second Envirofit stove. Because common cooking practices in the area require two simultaneous cooking pots (for example rice and beans, or *matooke* and some type of sauce), and the Envirofit is sized for one cooking pot, we gave a second Envirofit to permit normal cooking using only fuel-efficient stoves. We then collected our endline data, the core data we use in this study.

We tracked stove temperatures for approximately six months (April–September 2012). To track usage, we used small, inexpensive and unobtrusive sensors: stove use monitors (SUMs) that record stove temperatures without the need for an observer to be present.[3] Using SUMs to log stove temperatures was initially suggested by Ruiz-Mercado et al. (2008) and has been used successfully in various settings (Mukhopadhyay et al. 2012; Ruiz-Mercado et al. 2013; Pillarisetti et al. 2014). We installed SUMs on two Envirofits and two three-stone fires (by the end of the study numerous SUMs had been lost or burned up; therefore, at the end line we measured both Envirofits and the primary three stone fire).

We also performed standard kitchen performance tests (KPT) (Bailis, Smith, and Edwards 2007) in each household to measure the quantity of fuel wood used, record detailed food diaries, and measure household air pollution. The KPT lasts approximately a week and involves daily visits by a small team of researchers weighing wood, monitoring household air particulate monitors, and collecting survey data on stove usage over the last 24 hours and related topics.

---

[3] 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 which can be affixed to any stove type. Our SUMs record temperatures with an accuracy of +/- 1.3 degrees C up to 85°C. For additional details see the product description website at: http://berkeleyair.com/services/stove-use-monitoring-system-sums/ The SUMs cost approximately USD$16 each and could record temperature data for 24 hours a day for six weeks in a household before needing minimal servicing from a technician to download the data. After the data download they can be reset and re-used.

Comparing stove usage calculated from the temperature data collected by the SUMs in the week while KPT measurement teams are present versus stove usage in the week before and after the measurement week provides our test of a Hawthorne effect.

Throughout the study, field staff recorded about 2,400 visual observations of whether a stove was in use (on/off) when they visited homes to exchange stove usage monitors or gather data for the KPT. 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 1.7 million temperature readings. This process, detailed in Simons et al. (2014), allowed us to unobtrusively and inexpensively track daily stove usage on a large sample of households for six continuous months.[4]

Specification

Assign the subscripts $t=-1$ to the week prior to measurement week, $t=0$ to the measurement week, and $t=1$ to the week after the measurement week. Let the coefficient on stove type s = *TSF* for three-stone fire or *ENV* for Envirofit, and

[4] Overnight, while most participants report sleeping, SUMs record the residual heat absorbed in the large stones of the three stone fires and/or from coals banked overnight. Therefore our algorithm overestimates overnight cooking of three stone fires. We adjust for this in the subsequent analysis. For further discussion and a description of the technical adjustment see Harrell et al. (2016).

$Adjacent\_Week$ be a dummy variable for an adjacent week (*t=-1* or *t*=1). The regression is modeled using Ordinary Least Squares (OLS) as:

$$H_{it}^{S} = B^{S} * Adjacent\_Week + I_i + e_{it} \quad (1)$$

where H_{it}^{S} is the total hours cooked per day on stove type *s* for household *i* during the week, I_i is fixed effects for the individual household (which controls for household level characteristics that don't change over these three weeks like family size, income, housing, etc.), and e_{it} is an error term. The coefficient B^{S} is the estimate of how different (in hours cooked per day) the average adjacent week is compared to a measurement week on stove type *s*. Standard errors are clustered at the household.

To test the weeks separately, we use a slightly different specification. Let H_{it-1}^{S} be a dummy variable equal to 1 for the week before the measurement week (when *t=-1*) and 0 otherwise, and let H_{it+1}^{S} be a dummy variable equal to 1 for the week after the measurement week (when *t=1*) and 0 otherwise. Then the regression is modeled using OLS as:

$$H_{it}^{S} = \gamma_1^{S} * H_{it-1}^{S} + \gamma_2^{S} * H_{it+1}^{S} + I_i + e_{it} \quad (2)$$

where I_i is household fixed effects and γ_1^{S} is the estimate of the difference (in hours cooked per day) of the week before the measurement week compared to the measurement week. The coefficient of γ_2^{S} is the estimate of the difference cooked in the week after the measurement week compared to the measurement week. Standard errors are clustered at the household.

Results

In the week before the observers arrived (when *t=-1*), primary three-stone fires were used an average of 5.99 hours per day (95% CI = [4.77 to 7.21]) and combined usage on Envirofits was an average of 5.53 hours per day (95% CI = [4.36 to 6.71]). We first estimate equation 1, where we constrain the effect of the observers arriving to be the same magnitude (but opposite sign) as the effect of the observers leaving. On average, usage of the Envirofit stoves is 2.97 hours higher during the measurement week than during the adjacent weeks (95% CI = [1.79 to 4.15], $p<0.01$, Table 1, column 3). This increase is matched by a reduction of 1.78 hours in usage of the three-stone fire (95% CI = [0.86 to 2.70], $p<0.01$, col. 1).

In columns 2 and 4 we relax the assumption that stove usage is the same in the week prior to and the week after our measurement period. Contrasted with the measurement week, households use their primary three-stone fire 1.17 hours per day more in the prior week (95% CI = [0.10 to 2.24], $p<0.05$, col. 2) and 2.37 hours more in the following week (95% CI = [1.12 to 3.62], $p<0.01$). These coefficients are jointly significantly different than zero ($p<0.01$), but not statistically significantly different from each other ($p=0.10$).

The total usage of Envirofits follows a mirror image (col. 4), and is 2.58 hours per day lower in the week prior to measurement week than in measurement week (95%

CI = [1.21 to 3.94], $p<0.01$) and 3.30 hours per day lower the following week (95% CI = [2.04 to 4.57], $p<0.01$). These coefficients are jointly significantly different from zero ($p<0.01$), but not statistically significantly different from each other ($p=0.20$).

Adjusting for the Hawthorne effect

Because the kitchen performance test is widely used to measure the effects of new cookstoves on fuel usage and household air pollution (K. Smith et al. 2007; Berrueta, Edwards, and Masera 2008; Johnson et al. 2010)—as well as the basis for the measurement of carbon emissions—estimates of how new stoves affect fuel use and carbon emissions may be substantially biased. The same bias can arise in studies, such as ours, that also measure household air pollution or health effects with repeated household visits. We develop a basic framework for testing for the magnitude of this bias and examine its extent in our setting.

Basic framework

The field of epidemiology has "efficacy trials" that test the effects of an intervention under ideal conditions and "effectiveness trials" that test the effects of an intervention under realistic conditions (Flay 1986). In the context of cookstoves, the kitchen performance test provides a valid measure of how the new stove affects wood usage during the measurement week (as in an efficacy trial); however, we need to adjust for the gap in usage between measurement weeks and weeks when no observers are influencing behaviors to generalize to weeks without daily visits (that is, to estimate effectiveness). Next we consider various illustrative examples using data from our setting.

Illustrative examples

Table 2 presents the daily mean values of firewood consumption, particulate matter concentration and total three stone fire usage prior to the introduction of fuel-efficient stoves. The average household consumes 9.0 kgs of firewood per day (col. 1), has a daily concentration of PM2.5 of 428 $\mu g/m^3$ (col. 2) and cooks for a total of 14.0 daily hours (col. 3) across two three stone fires. To examine the bias introduced by the Hawthorne effect in our setting we need to know the expected biomass and pollution reductions for the new stove. To find the expected reduction we examine the "Emission and Performance Report" for the Envirofit G3300 performed by the Engines and Energy Conversion Lab at Colorado State University. These emissions measurements are based on accepted biomass stove testing protocols in a carefully monitored laboratory setting. The report (Figure 1) finds average improvements of 50.1% for fuel use and 51.2% for particulate matter emissions using the Envirofit G3300 versus a three stone fire (Envirofit Inc. 2011).

Using these mean values, we construct illustrative efficacy and effectiveness trails according to the framework above. For the purpose of our illustrative example, we assume a similar sized Hawthorne effect on the usage of the secondary three stone fires as well as what was observed on the first three stone fire (recall that attrition of sensors led us to measure fewer stoves in the endline).

Using the assumptions above, firewood consumed and daily PM2.5 concentrations were 16% lower when observers were present (as in an efficacy trial) than when they were not (as in an effectiveness trial, Table 3).

Bias introduced by Hawthorne effect

Table 4 presents a comparison of the endline to the baseline levels of daily cooking hours (on all stoves combined), daily firewood usage, and PM2.5 daily concentrations. Recall that at baseline no homes had any Envirofits, and at endline homes had two Envirofits. These results are not causal estimates, as seasonal or time effects may influence them.

When we use time periods when observers were present, between baseline and endline: cooking time rose 20% (from 14.0 to 16.8 hours),[5] firewood use declined 11% (9.0 to 8.0 kg/day), and particulate matter also fell 11% (from 428 to 382 μg/m^3). When examining weeks when observers were not present (as in an effectiveness trial), some important results are reversed. Now cooking time rose 24% (from 14.0 to 17.3 hours), firewood use rose 4% (from 9.0 to 9.3 kg/day) and exposure to particulate matter grew 4% (from 428 to 445 μg/m^3). That is, adjusting for the Hawthorne effect turned a decline of about 11% in wood use and participate matter into a small increase of about 4%.

[5] While total time cooking increases this is calculated over four stoves (two three stone fires and two Envirofit stoves) during the end line data collection period, but calculated over only two three stone fires during the baseline period. So it is likely that cooks actually spend less of their time cooking at end line because they have more stoves per meal at their disposal.

This illustrative example shows how important it is to account for Hawthorne effects in impact evaluations. Using the sample means from our data, and the emissions and performance report for the Envirofit G3300 the Hawthorne effect not only biases the magnitude of the change, but (with these assumptions) also reverses the direction of the change over time.

Conclusion

We demonstrate a technique to measure the magnitude of—and adjust for—a Hawthorne effect in a field experiment in the developing world. Given the push for more experiments in environmental economics (Greenstone and Gayer 2009), developing techniques to generate data that does not suffer from observer bias is necessary if the evaluations are to help make unbiased policy recommendations. In our specific setting, the findings of a large Hawthorne effect have implications for the impact of fuel-efficient stoves on fuel use and air particulates. The kitchen performance test is the current "gold standard" for generating Certified Emission Reductions that can be sold into the emissions trading markets of the Clean Development Mechanism. Our findings potentially call into question the veracity of these CO_2 reductions. More broadly, our results reinforce the importance for observed behaviors to be independently verified with unobtrusive monitoring.

While other forms of unobtrusive objective monitoring exist—such as using administrative records when reliable (Angrist, Bettinger, and Kremer 2006) or tracking take-up at a remote location via redeemed vouchers (Dupas 2009; Dupas 2014)—the recent explosion of small, inexpensive, and unobtrusive sensors expands researchers' ability to quantify and remove observation bias. A wide variety of emerging technologies can be utilized, a partial list includes: smart phones tracking locations through GPS, remote sensors that detect latrine usage (Clasen et al. 2012), sensors to remotely detect the use of water filters (Thomas et al. 2013), medical devices to monitor the hand hygiene of medical professionals (Boyce 2011), smart grid or other energy monitors (Darby 2010), and pedometers or other devices that monitor physical activity (Bravata et al. 2007). Adjusting for Hawthorne effects is essential if the results of impact evaluations are intended to generalize beyond periods of intense in-person observation.

Bibliography

Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2006. "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *American Economic Review* 96 (3): 847–62.

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

Banerjee, Abhijit, and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics* 1 (1): 151–78. doi:10.1146/annurev.economics.050708.143235.

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

Bento, Antonio, Daniel Kaffine, Kevin Roth, and Matthew Zaragoza-Watkins. 2014. "The Effects of Regulation in the Presence of Multiple Unpriced Externalities: Evidence from the Transportation Sector." *American Economic Journal: Economic Policy* 6 (3): 1–29. doi:10.1257/pol.6.3.1.

Berrueta, Víctor M., Rufus D. Edwards, and Omar R. Masera. 2008. "Energy Performance of Wood-Burning Cookstoves in Michoacan, Mexico." *Renewable Energy* 33 (5): 859–70. doi:10.1016/j.renene.2007.04.016.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data." *American Economic Review* 91 (2): 67–72.

Boyce, John M. 2011. "Measuring Healthcare Worker Hand Hygiene Activity: Current Practices and Emerging Technologies." *Infection Control and Hospital Epidemiology* 32 (10): 1016–28. doi:10.1086/662015.

Bravata, Dena M, Crystal Smith-Spangler, Vandana Sundaram, Allison L Gienger, Nancy Lin, Robyn Lewis, Christopher D Stave, Ingram Olkin, and John R Sirad. 2007. "Using Pedometers to Increase Physical Activity: A Systematic Review." *Jama* 298 (19).

Buchanan, Kathryn, Riccardo Russo, and Ben Anderson. 2014. "Feeding Back about Eco-Feedback: How Do Consumers Use and Respond to Energy Monitors?" *Energy Policy* 73 (June): 138–46. doi:10.1016/j.enpol.2014.05.008.

Clasen, Thomas, Douglas Fabini, Sophie Boisson, Jay Taneja, Joshua Song, Elisabeth Aichinger, Anthony Bui, et al. 2012. "Making Sanitation Count: Developing and Testing a Device for Assessing Latrine Use in Low-Income Settings." *Environmental Science & Technology* 46 (6): 3295–3303. doi:10.1021/es2036702.

Darby, Sarah. 2010. "Smart Metering: What Potential for Householder Engagement?" *Building Research & Information* 38 (5): 442–57. doi:10.1080/09613218.2010.492660.

Das, Jishnu, Jeffrey Hammer, and Kenneth Leonard. 2008. "The Quality of Medical Advice in Low-Income Countries." *Journal of Economic Perspectives* 22 (2): 93–114.

Duflo, Esther, R Glennerster, and Michael Kremer. 2008. "Using Randomization in Development Economics Research: A Toolkit." Edited by T Paul Schultz and John A Strauss. *Handbook of Development Economics* 4 (07): 3895–3962. doi:10.1016/S1573-4471(07)04061-2.

Dupas, Pascaline. 2009. "What Matters (and What Does Not) in Households' Decision to Invest in Malaria Prevention?" *American Economic Review* 99 (2): 224–30. doi:10.1257/aer.99.2.224.

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

Envirofit Inc. 2011. *G-3300 Woodstove Features*. http://www.envirofit.org/products/?sub=cookstoves&pid=10.

Ermes, Miikka, Juha Pärkkä, Jani Mäntyjärvi, and Ilkka Korhonen. 2008. "Detection of Daily Activities and Sports With Wearable Sensors in Controlled and Uncontrolled Conditions." *IEEE Transactions on Information Technology in Biomedicine* 12 (1): 20–26. doi:10.1109/TITB.2007.899496.

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.

Falk, Armin, and James J Heckman. 2009. "Lab Experiments Are a Major Source of Knowledge in the Social Sciences." *Science* 326 (5952): 535–38. doi:10.1126/science.1168244.

Flay, Brian R. 1986. "Efficacy and Effectiveness Trials (and Other Phases of Research) in the Development of Health Promotion Programs." *Preventive Medicine* 15 (5): 451–74.

Greenstone, Michael, and Ted Gayer. 2009. "Quasi-Experimental and Experimental Approaches to Environmental Economics." *Journal of Environmental Economics and Management* 57 (1): 21–44. doi:10.1016/j.jeem.2008.02.004.

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* forthcoming.

Johnson, Michael, Rufus Edwards, Victor Berrueta, and Omar Masera. 2010. "New Approaches to Performance Testing of Improved Cookstoves." *Environmental Science and Technology* 44 (1): 368–74. doi:10.1021/es9013294.

Leonard, Kenneth L. 2008. "Is Patient Satisfaction Sensitive to Changes in the Quality of Care? An Exploitation of the Hawthorne Effect." *Journal of Health Economics* 27 (2): 444–59. doi:10.1016/j.jhealeco.2007.07.004.

Leonard, Kenneth L, and Melkiory C Masatu. 2006. "Outpatient Process Quality Evaluation and the Hawthorne Effect." *Social Science & Medicine* 63 (9): 2330–40. doi:10.1016/j.socscimed.2006.06.003.

Leonard, Kenneth L., and Melkiory C. Masatu. 2010. "Using the Hawthorne Effect to Examine the Gap between a Doctor's Best Possible Practice and Actual Performance." *Journal of Development Economics* 93 (2): 226–34. doi:10.1016/j.jdeveco.2009.11.001.

Levine, David I, Theresa Beltramo, Garrick Blalock, Carolyn Cotterman, and Andrew M. Simons. 2016. "What Impedes Efficient Adoption of Products? Evidence from Randomized Sales Offers for Fuel-Efficient Cookstoves in Uganda." Berkeley, CA. http://escholarship.org/uc/item/2cb9g10z.

Levitt, Steven D, and John A List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives* 21 (2): 153–74.

———. 2011. "Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments." *American Economic Journal: Applied Economics* 3 (1): 224–38.

Miller, Grant, and A. Mushfiq Mobarak. 2013. "Gender Differences in Preferences, Intra-Household Externalities, and Low Demand for Improved Cookstoves." National Bureau of Economic Research. http://www.nber.org/papers/w18964.

Mueller, Florian, Frank Vetere, Martin R Gibbs, Stefan Agamanolis, and Jennifer Sheridan. 2010. "Jogging over a Distance: The Influence of Design in Parallel Exertion Games." *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, 63–68.

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.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2010. "The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India." *Economic Journal* 120 (546): F187–203. doi:10.1111/j.1468-0297.2010.02373.x.

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.

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.

Schwartz, Daniel, Baruch Fischhoff, Tamar Krishnamurti, and Fallaw Sowell. 2013. "The Hawthorne Effect and Energy Awareness." *Proceedings of the National Academy of Sciences of the United States of America* 110 (38): 15242–46. doi:10.1073/pnas.1301687110/.

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.

Smith, Eleanor. 2014. "Better Hygiene Through Humiliation." *The Atlantic*, August 13.

Smith, Kirk, Karabi Dutta, Chaya Chengappa, P.P.S. Gusain, Omar Masera, Victor Berrueta, Rufus Edwards, Rob Bailis, and Kyra Naumoff Shields. 2007. "Monitoring and Evaluation of Improved Biomass Cookstove Programs for Indoor Air Quality and Stove Performance: Conclusions from the Household Energy and Health Project." *Energy for Sustainable Development* 11 (2): 5–18.

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.

Srigley, Jocelyn A, Colin D Furness, G Ross Baker, and Michael Gardam. 2014. "Quantification of the Hawthorne Effect in Hand Hygiene Compliance Monitoring Using an Electronic Monitoring System: A Retrospective Cohort Study." *BMJ Quality & Safety* 23 (12): 974–80. doi:10.1136/bmjqs-2014-003080.

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.

Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer, et al. 2011. "Being Surveyed Can Change Later Behavior and Related Parameter Estimates." *Proceedings of the National Academy of Sciences of the United States of America* 108 (5): 1821–26. doi:10.1073/pnas.1000776108.

Table 1

Regressions testing for Hawthorne effect: estimates of effects of the presence of measurement team in primary three stone fire (TSF) usage and combined Envirofit usage, the coefficients represent the change in hours cooked per day compared to hours cooked per day in the measurement week

	Primary TSF		Combined Envirofit	
	(1)	(2)	(3)	(4)
Week prior to and after measurement week constrained to be equal	1.78***		-2.97***	
	(0.46)		(0.60)	
Week prior to measurement week		1.17**		-2.58***
		(0.54)		(0.69)
Week after measurement week		2.37***		-3.30***
		(0.63)		(0.64)
Household fixed effects	Yes	Yes	Yes	Yes
Observations	316	316	229	229
R-squared	0.82	0.82	0.79	0.79
Household clusters	118	118	89	89

Standard errors clustered at household level in parentheses

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

Note: The unit of analysis is a measurement "week" (approximately 72 hours) at a household. The specification in columns 1 and 3 imposes that the weeks prior to and after the measurement week are equal. The specification in columns 2 and 4 tests usage in the week prior to and after the measurement week separately. The coefficient estimates in column 2 are jointly significantly not equal to zero ($p<0.01$), but not statistically different from each other ($p=0.10$). The coefficient estimates in column 4 are jointly significantly not equal to zero ($p<0.01$), but not statistically different from each other ($p=0.20$).

Table 2

Daily mean firewood consumption, particulate matter and three stone fire usage prior to introduction of fuel-efficient stoves

	Wood Consumed (kgs)	PM2.5 (µg/m^3)	Three Stone Fire (hours)
	(1)	(2)	(3)
Mean Values	8.98***	427.79***	13.95***
	(0.33)	(23.18)	(1.03)
Observations	568	609	339
Household clusters	160	159	102

Standard errors clustered at household level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Columns 1, 2, and 3 present the average daily wood consuption, average daily PM2.5 reading, and the total hours of combined cooking on two three stone fires per household prior to receiving a fuel-efficient stove, respectively. Observations are at the household-day level.

Table 3

Estimates of biomass usage and indoor air pollution with and without in-person observers

Efficacy Trial (effects of an intervention during week with observers)			
	Daily hours	Biomass per hour	Total biomass (kg)
Total three stone fire usage	8.30	0.64	5.31
Total Envirofit usage	8.50	0.32	2.72
Totals:	16.80		8.03
	Daily hours	PM2.5 per hour	Total PM2.5 (µg/m^3)
Total three stone fire usage	8.30	30.67	254.56
Total Envirofit usage	8.50	14.97	127.25
Totals:	16.80		381.81
Effectiveness Trial (effects of an intervention during week without observers)			
	Daily hours	Biomass per hour	Total biomass (kg)
Total three stone fire usage	11.81	0.64	7.56
Total Envirofit usage	5.53	0.32	1.77
Totals:	17.34		9.33
	Daily hours	PM2.5 per hour	Total PM2.5 (µg/m^3)
Total three stone fire usage	11.81	30.67	362.21
Total Envirofit usage	5.53	14.97	82.78
Totals:	17.34		444.99

Note: Daily hours for the effectiveness trial are taken from the data for the week prior to the KPT. Recall that only about one fifth of the secondary three stone fires had iButtons on them at this point in our experiment. For the purpose of this illustrative table we make the assumption that households with missing values for the secondary three stone fire are equal to the mean value observed for the one fifth of the sample that had an hourly usage reading for the secondary three stone fire. Daily hours for the efficacy trial are based on the Hawthorne effects presented in Table 1. The consumption rates for biomass and PM2.5 with the three stone fires are calculated prior to the introduction of fuel-efficient stoves using the values in Table 2 (8.98 kg/13.95 hours = 0.64 kgs/hour and 427.79 µg/m^3/13.95 hours = 30.67 µg/m^3/hour). The consumption rates for the Envirofit G3300 are calculated using the emissions testing report in Figure 1 (0.64 kgs/hour * 0.499 = 0.32 kgs/hour and 30.67 µg/m^3/hour * 0.488 = 14.97 µg/m^3/hour).

Table 4

Bias introduced by the Hawthorne effect

	Daily cooking (hours)	Total biomass (kg)	Total PM2.5 ($\mu g/m^3$)
Baseline	13.95	8.98	427.79
Efficacy (observers present)	16.80	8.03	381.81
Effectiveness (no observers)	17.34	9.33	444.99

Note: These calculations are illustrative based on the mean values of data collected in the field and the emissions and performance report performed in a laboratory. These calculations assume a similarly sized Hawthorne effect on the secondary three stone fire as what we observed on the primary three stone fire.

Figure 1

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

Note: The report can be downloaded at http://www.envirofit.org/images/products/pdf/g3300/G3300Cert.pdf

OTHER A.E.M. WORKING PAPERS

WP No	Title	Fee (if applicable)	Author(s)
2016-07	Economics and Economic Policy		R. Kanbur
2016-06	W. Arthur Lewis and the Roots of Ghanaian Economic Policy		R. Kanbur
2016-05	Capability, Opportunity, Outcome - -and Equality		R. Kanbur
2016-04	Assessment of New York's Pollution Discharge Elimination Permits for CAFO's: A Regional Analysis		Enahoro, D., Schmit T. and R. Boisvert
2016-03	Enforcement Matters: The Effective Regulation of Labor		Ronconi, L. and R. Kanbur
2016-02	Of Consequentialism, its Critics, and the Critics of its Critics		Kanbur. R
2016-01	Specification of spatial-dynamic externalities and implications for strategic behavior in disease control		Atallah, S., Gomez, M. and J. Conrad
2015-13	Networked Leaders in the Shadow of the Market - A Chinese Experiment in Allocating Land Conversion Rights		Chau, N., Qin, Y. and W. Zhang
2015-12	The Impact of Irrigation Restrictions on Cropland Values in Nebraska		Savage, J. and J. Ifft
2015-11	The Distinct Economic Effects of the Ethanol Blend Wall, RIN Prices and Ethanol Price Premium due to the RFS		de Gorter, H. and D. Drabik
2015-10	Minimum Wages in Sub-Saharan Africa: A Primer		Bhorat, H., Kanbur, R. and B. Stanwix
2015-09	Optimal Taxation and Public Provision for Poverty Reduction		Kanbur, R., Pirttilä, J., Tuomala, M. and T. Ylinen
2015-08	Management Areas and Fixed Costs in the Economics of Water Quality Trading		Zhao, T., Poe, G. and R. Boisvert
2015-07	Food Waste: The Role of Date Labels, Package Size, and Product Category		Wilson, N., Rickard, B., Saputo, R. and S. Ho
2015-06	Education for Climate Justice		Kanbur, R.

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