# Evaluating the effectiveness of household energy interventions in rural Senegal using experimental and quasi-experimental methods

Pre-analysis plans

October 2018

Principal research team:

Marc Jeuland, Duke University; Luciane Lenz, RWI; Erin Litzow, Duke University; Samba Mbaye, Université Gaston Berger; Ousmane Ndiaye, Université Gaston Berger; Jörg Peters, RWI; Maximiliane Sievert, RWI; Faraz Usmani, Duke University

Note:

This document outlines the pre-analysis plan for one out of three distinct studies associated with the comprehensive 3ie-funded impact evaluation of current and proposed rural-energy interventions in Senegal. It outlines an experimental evaluation of the welfare impacts associated with the adoption and use of two types of improved cookstoves (ICS) as compared to traditional stoves. The pre-analysis plan broadly follows the "checklist" on pre-analysis plans suggested by Mckenzie (2012).

# Impacts of improved and clean cookstoves on rural Senegalese households

Date prepared: 13 October 2018

Study location and timeline: Senegal; January 2018 to April 2019

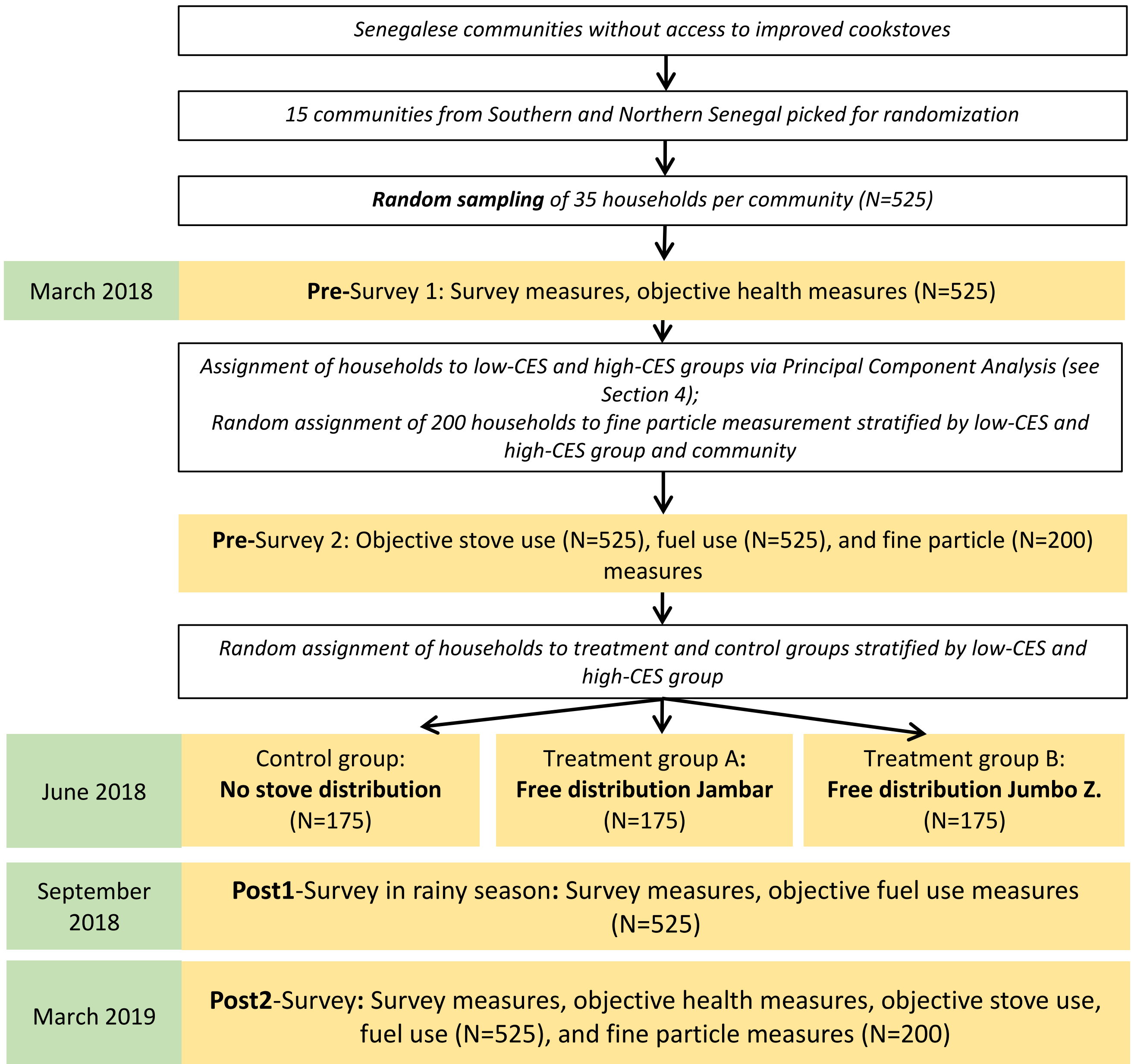
**1. Description of the sample to be used in this study**

Our sample will consist of 525 households in 15 rural Senegalese communities. Figure 1 describes the sampling design. The 15 communities are to be located in two regions: the first in northern Senegal (Saint-Louis or aLouga or Linguère) and the second in southern Senegal (Kaffrine or Fatick or Kolda). Within communities, households will be selected for invitation to participate in the baseline survey from village lists. Based on data collected during the baseline survey, households will then be assigned to one of three groups via random stratified sampling. The three groups will consist of a control group, treatment group A and treatment group B, and the key stratification variable will be a PCA-based categorization (low-CES and high-CES) that is motivated by some of the elements included in the *Cooking Energy System (CES)* concept developed by Energising Development (Endev), an energy access partnership currently financed by six donor countries: the Netherlands, Germany, Norway, United Kingdom, Switzerland and Sweden (GIZ 2017).

The *Cooking Energy System (CES)* is based on a hypothesis that some households have "cleaner" or more advanced systems than others, across a range of variables that include fuels, stove types, ventilation conditions, type of food preparation, behavior of the cook and other household members, and other factors that influence both household members' health and household fuelwood consumption. Our stratification was based primarily on the ventilation and technology elements of this system, although it also accounts for the number of cooks and stoves used by the household (see further details below).

Shortly after the baseline data collection, treatment group A households will receive a Jambar Jaboot woodfuel stove free of charge, treatment group B households will receive a Jumbo Zama free of charge, and control group households will receive a wax print as compensation for participation. We will revisit all households twice, once for a midline survey, and once for a more intensive follow-up survey.

*Figure 1: Sampling Design*




## 2 Key data sources

Data for the study will come from three different surveys (Pre, Post1, Post2). Table 1 summarizes the data that will be collected by survey.

*Table 1: Data and data sources*

| Pre-Survey | *Post1-Survey* | *Post2-Survey* | |
|---|---|---|---|
| x | *x* | *x* | <u>**Survey measures:**</u> Household demographics and composition; environmental quality perceptions; knowledge of cooking technologies; measures of fuel and stove use; measures of health; cooking and solid fuel collection practices and time use; psychosocial preferences (e.g. risk aversion); socioeconomic status; and contextual differences. |
| x | | x | Objective **health measures** from the household's primary cook. |
| x | | x | Objective, sensor-based measures (SUMs) of **stove use.** |
| x | x | x | Objective measures of **fuel use.** |
| x | | x | Objective measures of **fine particles (PM2.5) exposure** of a subsample of households' main cooks, and **fine particle concentration** in a subsample of households' kitchens. |

## 3 Hypotheses to be tested through the causal chain

a. $H_0/H_a$: No impact (positive impact) of receiving *any* treatment (Treatment A or Treatment B) mainly on

   i. Objective and subjective measures of fuel use (pre/post1/post2 periods)

   ii. Objective and subjective stove use measures (pre/post2 only for objective measures, pre/post1/post2 for subjective measures)

   iii. Objective and subjective health measures(pre/post2 only for objective measures, pre/post1/post2 for subjective measures)

b. Effects by treatment arm:

   i. $H_0/H_a$: No impact (positive impact) of receiving Treatment A on indicators listed in 3a (for the relevant periods).

   ii. $H_0/H_a$: No impact (positive impact) of receiving Treatment B on indicators listed in 3a (for the relevant periods).

c. Heterogeneity analysis according to the CES index: We will construct this index by applying principal components analysis to the baseline data for ventilation-related variables appearing in the cooking energy system framework (see Section 4). We will stratify the sample into two groups according to the first principal component of this analysis (low and high), and test for heterogeneity in the effects across strata.

   i. $H_0/H_a$: No difference (difference) between the impacts of receiving *any* treatment for households ranked as low CES compared to households ranked as high CES (for the relevant periods).

   ii. $H_0/H_a$: No difference (difference) between the impacts of receiving Treatment A for households ranked as low CES compared to households ranked as high CES (for the relevant periods).

   iii. $H_0/H_a$: No difference (difference) between the impacts of receiving Treatment B for households ranked as low CES compared to households ranked as high CES (for the relevant periods).

d. For a limited set of variables (see Table 1) we conduct an analysis for heterogeneous treatment effects over time, i.e. for the 6-month period following treatment and the 12-month period following treatment. This includes objective fuel use, self-reported respiratory illness and eye irritation, measures of reported fuel and stove use, and time spent collecting fuel and cooking. Variation may result from seasonality changes or usage changes over time.
   i. $H_0/H_a$: No difference (difference) between the impacts of receiving *any* treatment six months after the treatment compared to 12 months after the treatment.
   ii. $H_0/H_a$: No difference (difference) between the impacts of receiving Treatment A six months after the treatment compared to 12 months after the treatment.
   iii. $H_0/H_a$: No difference (difference) between the impacts of receiving Treatment B six months after the treatment compared to 12 months after the treatment.

These hypotheses will be tested using simple means comparisons, as well as difference-in-difference regression estimates, also controlling for demographic and socio-economic differences across households. Furthermore, the three survey waves for a subset of variables allow for analyzing the variation in treatment effect over time.

## 4 How main variables will be elicited and constructed

The main outcome variables as introduced above include the following:

a. Objective and subjective measures of fuel use focus on consumption of fuel, financial expenses and time use (see Brooks et al. 2015). The relevant variables will likely include:
   i. Subjective firewood use (kg/day)
   ii. Subjective fuel collection time (hr/wk)
   iii. Money spent on fuel (CFA/mo)
   iv. Weighed measures of fuel use (kg/day)

b. Objective and subjective stove use measures
   The main subjective variables to capture fuel use (Brooks et al. 2015) are clean stove use (min/day) and traditional stove use (min/day)
   In addition, our objective measure is a sensor-based (SUMS; from Berkeley Air) indicator of stove use (see Ruiz-Mercado et al. 2008).

c. Objective and subjective health measures
   The objective variables to measure the main cook's health include measurements of systolic and diastolic blood pressure, pulse oximetry, and biomarkers of inflammation (namely CRP, or C-Reactive protein). The main subjective variables are self-reported prevalence of respiratory illness and eye diseases (Lewis et al. 2015; Baumgartner et al. 2011) of the main cook and household members aged below 10 or above 59 years.

We conduct a Principal Component Analysis based on eight variables that are prominent in Energising Development's (EnDev) proxy-indicator approach for assessing the quality of a CES[1]. Based on the resulting PCA score, we assign households to either a low-PCA category or a high-PCA category. The PCA categories are used for random stratification (see Figure 1) and heterogeneity analysis. The eight variables for PCA analysis are listed in the following.

i. Dummy equals 1 if main kitchen is indoor with at least an overhanging roof, and 0 if cooks outdoor.
ii. Kitchen location
    - 1= Inside the building without seperation
    - 2= Inside the building without seperation
    - 3= Inside the building in separate room
    - 4= Room attached to building with separate entrance, or independent building
    - 5= Outside cooking
iii. Kitchen ventilation
    - 1= No opening except for the door
    - 2= Openings below the height of the door
    - 3= Open air or substantial openings in the roof or above the height of the door
iv. Kitchen volume
v. Dummy equals 1 if uses charcoal or gas for cooking
vi. Dummy equals 1 if uses multiple stoves
vii. Cooking time (min/day)
viii. Number of primary cooks within household

## 5 Sample size and power calculations

In the following, we provide power calculations for key outcome variables. We have not done power calculations related to the biomarkers due to a lack of prior data on the likely levels of inflammation and exposure in this population. For the main analysis of the outcomes discussed above except perhaps respiratory illness, we are confident that the detectable differences will be observed as long as households use the ICS. Power for the heterogeneity analysis is much less certain, due to the lack of data and research pertaining to non-stove aspects of the CES. Changes in blood pressure will be used to indicate reduced exposure (Baumgartner et al. 2014).

Our sample size calculations focus on a number of outcomes measured in the study's impact arm, for which we can draw on prior data obtained in other surveys from similar locations in Senegal (Bensch & Peters 2015). The calculations are based on the following assumptions:

a. Treatment assignment (with one of the two improved biomass stoves) will be randomized at the household level (no need to consider inter-cluster correlation);

[1] For an introduction see https://endev.info/images/f/f7/Factsheet_EnDev_CES_EN.pdf

b. Baseline use of clean cooking options will be low or negligible in this rural sample (on average less than 10 minutes per household-day);
c. A sample size of 175 households per treatment arm (or 525 overall including control households), 400 of whom (from the impact and control arms) will be monitored using biomarkers; and
d. Health outcomes materialize at the individual level, while the treatment is assigned at household level. Thus, we have to account for correlation at the household level. The correction factor very much depends on the within group variation. Here, we assume that health outcomes are perfectly correlated between members of one household, which is very conservative from a power perspective.

The key results are summarized in Table 1, where we detail our assumptions about baseline levels of key indicators, and the detectable difference across groups in the main analysis under varying assumptions about significance (α) and power (β).

*Table 2: Summary of power calculations for key outcomes*

| | | Detectable difference (α; β) | | |
|---|---|---|---|---|
| Outcome | Baseline mean (s.d.) | (0.1, 0.9) | (0.1, 0.8) | (0.05, 0.9) |
| Clean stove use (min/day) | 10 (45) | 199% | 168% | 221% |
| Trad. stove use (min/day) | 323 (173) | -24% | -20% | -26% |
| Firewood use (kg/day) | 13.0 (10.3) | -35% | -30% | -39% |
| Fuel collection time (hr/wk) | 14.3 (13.0) | -40% | -34% | -45% |
| Systolic blood pressure | 135 (20) | -7% | -6% | -7% |
| Diastolic blood pressure | 78.3 (12) | -7% | -6% | -8% |
| Respiratory illness prevalence | 21 % | -71% | -62% | -77% |

## 6 Specify the treatment effect equation to be estimated

To test the hypothesis outlined in **3a**, we will first apply the following specification:

$$Y_i = \alpha + \theta stove_i + \epsilon_i \qquad (1a),$$

where $Y_i$ represents the different outcomes of interest as outlined above (3a.i, 3a.ii, 3a.iii) for household $i$, measured after the intervention. Our coefficient of interest is θ, which represents the impact of being assigned to *any* of our two treatment arms, denoted by the binary variable $stove_i$. Finally, $\epsilon_i$ represents the unobserved household-specific residual.

We then test for differences between treatment A and treatment B (see **3b**) by modifying equation (1a) as follows:

$$Y_i = \alpha + \theta_1 stoveA_i + \theta_2 stoveB_i + \epsilon_i \qquad (2a),$$

where $stoveA_i$ and $stoveB_i$ represent binary variables for random assignment to stove A and stove B, respectively. Random assignment to the control group will be the omitted category.

To test the hypothesis outlined in **3c**, we next conduct a heterogeneity analyses that allows estimation of these impacts according to CES category (low or high). We here re-estimate the regression models from (1a) and (2a), separately for two sub-samples that represent households with a low and high CES score at baseline. We then compare the $stove$ coefficients across the models using formal hypothesis testing.

In expectation, the random assignment of our treatment should ensure statistical indistinguishability between the three groups concerning their observable and unobservable characteristics, such as socio-economic characteristics, household demographics and composition. We will control for imbalances between groups at baseline that occur by chance, however, and test for the robustness of results we obtained in (1a) and (2a) by including additional covariates, as illustrated in (1b) and (2b), respectively. We will make the same adjustments to the CES sub-sample estimations, further controlling there for other variables included in the EnDev framework.

$$Y_{ij} = \alpha + \theta stove_i + \mathbf{X}_i'\boldsymbol{\beta} + \gamma_j + \epsilon_{ij} \qquad (1b)$$

$$Y_{ij} = \alpha + \theta_1 stoveA_i + \theta_2 stoveB_i + \mathbf{X}_i'\boldsymbol{\beta} + \gamma_j + \epsilon_{ij}. \qquad (2b),$$

where $\mathbf{X}_{ij}'$ represents a series of household variables that represent household demographics, household composition, and socio-economic variables, such as age, educational attainment, household size and assets, and baseline measures of variables found to be unbalanced following randomization. Additionally, village fixed effects will be introduced, as represented by $\gamma_j$.

In addition, we will test the robustness of results from (1a) by re-estimating the treatment effects of the models outlined above with a standard difference-in-difference (DiD) estimation. The approach controls for time-invariant unobserved differences between groups. We re-test the hypotheses outlined in **3a** and **3b** as following:

$$Y_{it} = \alpha + \beta_1 Sto_i + \beta_2 Post_t + \delta_1 (Sto_i * Post_t) + \epsilon_{it} \qquad (1c),$$

where in $\delta_1$ is the DiD estimator for the impact of the intervention of a treated household $i$. $Sto$ denotes the stove treatment, i.e. it equals unity when any improved cookstove has been received, and $Post$ denotes a post intervention dummy. $\epsilon_i$ represents the unobserved household specific residual. The same specification will be applied to the CES sub-samples, where we compare the treatment effect sizes, i.e. the DiD coefficient, across sub-samples (see hypothesis **3c)**.

To then test for the effects by treatment arm A and B in a DiD set-up, we modify the specification from (2a) and estimate:

$$Y_{it} = \alpha + \beta_1 StoA_i + \beta_2 StoB_i + \beta_3 Post_t + \delta_1(StoA_i * Post_t) + \delta_2(StoB_i * Post_t) + \epsilon_{it} \qquad (2c),$$

where $\delta_1$ and $\delta_2$ are the DiD estimators for the effect of receiving Treatment A and Treatment B respectively, i.e. $StoA$ and $StoB$ denote dummies for having received the different stove treatments.

The same specification will again be applied to the CES sub-samples in order to compare the DiD coefficient across sub-samples (see hypothesis **3c)**.

Lastly, we test hypotheses **3d**. Here, we use the multiple (three) waves of the survey to conduct a difference-in-difference identification strategy with one pre-treatment and two post-treatment periods. Differences in effect sizes between the periods may reflect seasonal differences in treatment effects or changes in stove use over time. Note that it is for a limited set of variables only that we will have three points in time (namely all those covered in the midline, see 2.d). We modify equation (1c) as following:

$$Y_{it} = \alpha + \beta_1 Sto_i + \beta_2 Post1_t + \beta_3 Post2_t\ \delta_1(Sto_i * Post1_t) + \delta_2(Sto_i * Post2_t) + +\epsilon_{it} \qquad (1d),$$

Where $\delta_1$ and $\delta_2$ and the treatment effects by survey wave. $Post1$ and $Post2$ are post-treatment dummies which equal one for the first and second post-treatment period respectively. Analogous to equation (1d), we will re-estimate equation (2c) to test for heterogeneous effects by treatment arms A and B (not shown here).

## 7 Plan for how to deal with multiple outcomes and multiple hypothesis testing

We will employ established approaches (e.g. Holm, 1979 or the Benjamini-Hochberg (1995) False Discovery Rate correction) as well as more recent innovations (e.g. List et al., 2016) to adjust the inference associated with our multiple subgroup analyses. Specifically, we will apply the Anderson (2008) correction for hypothesis testing of outcomes that are in similar 'families', e.g., health outcomes, or time use outcomes.

## 8 Procedures to be used for addressing survey attrition and missing data

We do anticipate only a small risk of attrition over the course of our study, given that it covers a period of one year only, and that interaction with sample communities within this period is high (four household visits in total). Furthermore we provide incentives by distributing products at zero-cost (ICS and wax print), and we select rural areas in rather small communities and gather GPS data on household location. We will first evaluate the degree of attrition, and whether attrition is balanced across our treatment arms. For an overall rate of attrition less than or equal to five percent that is also balanced across treatment arms, we will exclude households that dropped out of our study from our analyses. For any deviation from these criteria, we will estimate inferential bounds for the size of our treatment effects, as described by Lee (2009). This semi-parametric approach relies on relatively weak assumptions about the how a randomly assigned treatment influences outcomes of interest to obtain intervals on the estimated size of the treatment effect in the presence of non-random attrition.

## 9 Procedures to be used for addressing outliers

We will deal with outliers by capping unbounded variables at the $95^{th}$ or $99^{th}$ percentile of the observed values in our data, as well as dropping affected observations, and testing sensitivity to these approaches by comparing the estimates for the coefficients of interest.

## 10 Procedures to be used for addressing missing covariate values

We will follow Lin, Green and Coppock (2016) in treating missing covariates. If no more than 10 percent of the covariate's values are missing, we will recode the missing values to the overall mean or village means (again testing sensitivity of estimates to these approaches by comparing results with those obtained from the sample with non-missing covariates). If more than 10 percent of the covariate's values are missing, we will include a missingness dummy as an additional covariate and recode missing values to 0.

## 11 Procedures to be used for addressing missing dependent variables

To deal with missing values on our outcome measures, we will adopt the approach described in Kling, Liebman and Katz (2007) and impute missing values by setting them equal to the mean of the respective outcome variable for the relevant treatment group, and testing sensitivity of main coefficient estimates to this approach by comparing results with those obtained from the sample with non-missing outcome variables.

## References

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American statistical Association*, 103(484), 1481-1495.

Baumgartner, J., J.J. Schauer, M. Ezzati, L. Lu, C. Cheng, J.A. Patz, and L.E. Bautista (2011). Indoor air pollution and blood pressure in adult women living in rural China. *Environmental health perspectives* 119(10).

Baumgartner, J., Y. Zhang, J.J. Schauer, W. Huang, Y. Wang, and M. Ezzati (2014). Highway proximity and black carbon from cookstoves as a risk factor for higher blood pressure in rural China. *Proceedings of the National Academy of Sciences* 111(36): 13229-13234.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society*. Series B (Methodological), 289-300.

Bensch, G., and J. Peters. (2012). "A Recipe for Success? Randomized Free Distribution of Improved Cooking Stoves in Senegal." *Ruhr Economic Papers* #325. http://ssrn.com/abstract=2030746.

Brooks, N., Bhojvaid, V., Jeuland, M., Lewis, J.J., Patange, O. & Pattanayak, S.K. (2015). How much do alternative cookstoves reduce biomass fuel use? Evidence from north India. *Resource and Energy Economics* 43: 153-171.

GIZ (2017). EnDev's proxy-indicator approach for assessing the quality of a Cooking Energy System.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, Theory and Applications*, *6*(2), 65–70.

Kling, Jeffrey R, Jeffrey B Liebman and Lawrence F Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica* 75(1):83–119.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies*, 76(3), 1071–1102.

Lewis, J., V. Bhojvaid, N. Brooks, I. Das, M.A. Jeuland, O. Patange, and S.K. Pattanayak. (2015). "Piloting Improved Cookstoves in India." *Journal of Health Communication*, Volume 20, Issue 1, Pages 28-42, http://www.tandfonline.com/doi/full/10.1080/10810730.2014.994243?scroll=top&needAccess=true.

Lin, Winston, Donald Green and Alexander Coppock. (2016). "Standard Operating Procedures: A

Safety Net for Pre-Analysis Plans." *PS: Political Science & Politics*, 49 (3), 495-500.

List, J. A., Shaikh, A. M., & Xu, Y. (2016). Multiple Hypothesis Testing in Experimental Economics. National Bureau of Economic Research (NBER) Working Paper № 21875.

Mckenzie, D. (2012). *A pre-analysis plan checklist*. *Development Impact*. Retrieved 13 September 2017, from https://blogs.worldbank.org/impactevaluations/a-pre-analysis-plan-checklist

Ruiz-Mercado, I., Lam, N. L., Canuz, E., Davila, G., & Smith, K. R. (2008). Low-cost temperature loggers as stove use monitors (SUMs). *Boiling Point*, 55, 16-18.