Increasing Household Electrification Rates in rural Uttar Pradesh with Connection Campaigns
A Pre-Analysis Plan

Brian Blankenship
Columbia University

Ryan Kennedy
University of Houston

Johannes Urpelainen*
Columbia University

Jason Chun Yu Wong
Columbia University

January 9, 2018

*Corresponding author. Address: 420 West 118th Street, 712 International Affairs Building. New York, NY 10027, USA. Tel: +1-734-757-0161. Email: ju2178@columbia.edu.
1 Motivation

This is a pre-analysis plan for a randomized controlled trial in Uttar Pradesh (UP), India. In this study, we are interested in testing whether an intervention to encourage households to set up legal electricity connections can make an impact on the total number of rural households with access to electricity. We survey 10 households in 200 habitations in the Sitapur district and the Bahraich district of Uttar Pradesh. We thus have a sample size of 2,000 respondents altogether.

The primary motivation of this project is to evaluate the capacity for social workers to encourage electrification as a method for meeting the government’s goal of 24x7 power to all households in UP by 2019. The electrification of villages in rural Uttar Pradesh has proceeded rapidly in recent years, but the level of consumption has remained well below the national average (India Ministry of Power, 2017). One of the largest hurdles to increasing consumption is that, even in electrified villages, many households still lack domestic connections. A large part of the success or failure of the government’s ambitious goal of providing 24x7 power to households by 2019 is increasing these household connection rates.

Our fieldwork in different districts of Uttar Pradesh – Hardoi, Sitapur, Unnao, and Mahmudabad – suggests that few people are aware of the procedures for obtaining a household connection. In January 2017, we visited ten villages in these districts to learn about people’s perceptions and found that none of the household heads without electricity at home knew how to apply for a connection. Most of them also overestimated the cost of a connection by a substantial amount: even though the real connection cost for a 1-kW connection was typically about 1,500 rupees, the majority of the people we interviewed thought it was over 2,500 rupees (Jain et al., 2015). This lack of knowledge and the overly pessimistic assessment of the cost of a connection suggest that an awareness campaign could allow major improvements in household electrification.

Increasing electricity access in rural areas could create opportunities for economic growth by allowing villagers to use power for economic activity. Recognizing this, the government of India initiated the large-scale “Saubhagya” scheme in late-2017 specifically to promote this type of last-mile connectivity to willing households. The cornerstone of this push is intensive efforts to encourage applications for legal electricity connections. The purpose of this project is to test whether these
project that help households with submitting applications for domestic connections could be an easy and inexpensive way to increase rural electrification rates in already electrified villages.

2 Proposed Approach

In the proposed project, enumerators hired for conducting the research play the role of a social worker to help households with applications as part of a “campaign” to increase electricity access. The enumerator meets with household heads to inform them about the cost and procedures of applying for a domestic connection. The meetings are an easy and inexpensive way to inform households about the application procedure.

The enumerator then collects names of households interested in applying for a connection. These households receive help from the enumerator with filling applications, supporting documents, and photographs. The enumerator also helps with transportation and accompanies households to the electricity substation. Thus, households can easily apply for connections, provided they are willing to pay the proper fee.

The effectiveness of this campaign can be tested by comparing electrification rates for households in habitations covered by the campaign to those not covered. If the approach is successful, we expect to see a rapid increase in domestic connections.

3 Sampling, Treatment, Randomization, and Data Collection

3.1 Sampling

Our sample consists of 2,000 non-electrified households in 200 habitations in the Sitapur district and the Bahraich district of Uttar Pradesh. These districts are located about 100-150 km from the regional capital of Lucknow. The electrification profile, generally adequate supply of electricity but low connection rates, works well for the purposes of our study.

Habitations need to meet the following requirements for selection into the sample:

- There must be an adequate supply of electricity (minimum 10 hours, ideally higher) to ensure that household electrification is possible at a level that would be desirable for households.

- There must be low electrification rates (below 25%) to ensure that there is still potential demand.
They must be at least 1km from other habitations in the sample.

We will select 200 habitations for the project. Enumerators are to interview 10 households in each habituation – participation is voluntary. One habituation can have at most one neighborhood and there will be a maximum one campaign per habituation.

The sampling and data collection will be conducted as follows:

1. Enumerators visit the study area to prepare a list of 200 habitations with low electrification rates, below 25% percent, and sufficient hours of electricity.

2. The list of 200 habitations is sent to the researchers for randomization at the habitation level (see below for details).

3. Enumerators conduct a baseline survey in all habitations.

4. In treatment habitations, the baseline survey is followed by a presentation on how to apply for electricity connections. Interested households are also given information about how to apply for an electricity connection and assistance with the process.

5. After the treatment is implemented, mobile phone surveys are used to collect the endline data. A small team of enumerators is sent back to the study area to contact any households unavailable through the mobile phone and to validate the quality of the mobile phone data through random spot checks.

Since the experiment will be conducted during a period where the government is implementing the Saubaghya scheme for electrification, the research team will make weekly calls to government officials to see whether any government-run Saubaghya camps have opened in the study area and will adjust accordingly.

3.2 Treatment

The treatment is a household visit that explains how the household can apply for an electricity connection and offers support with the application. The habitations are identified in advance by one Morsel enumerator (see above for requirements).
Households can participate if they are non-electrified, within 40 meters of a power pole, and express interest in electricity connection when asked. Potential participants will not be told about the procedure or cost of receiving a connection prior to the provision of application information if they are in the treatment group, and will not be provided this information at all in the control group. Households will be recruited using general language, asking if they are generally interested in a connection. The same recruiting language will be used for the treatment and control group.

Prior to going to the villages, the local power house is contacted to inquire about any steps that need to be taken for the campaign and acquire a memorandum of understanding, if needed.

In the campaign, the 10 participating households from the selected habitation go through the following steps – each individually:

- A baseline survey is conducted to collect initial attitudes for both the treatment and control.

- Those in the treatment condition are given a detailed introduction to electricity application: cost, procedure, timeline, and the support we are offering. This introduction only focuses on information that is relevant to the household (e.g. only Below Poverty Line (BPL) households, who are provided with free electricity connections, are told BPL requirements).\(^1\)

- Treated participants are reminded that all new connections are metered and that the fixed meter rent is 50 rupees per month.

- Cheat sheets (3 pages) are provided to treatment participants that give the relevant information and requirements.

- Treatment participants are also shown pictures of what is needed to apply for an electrical connection: application form, photos, housing information, ID (BPL if appropriate), and payment.

- Treatment participants make a decision of whether to submit an application for an electrical connection.

\(^1\) BPL households are defined according to the 2011 Socio Economic and Caste Census (SECC). Whether a household qualifies as BPL is based on an index system that takes into account items like household possessions, deprivation, and occupation.
Treatment participants fill out a survey about the campaign and the utility of the information provided.

After the campaign, enumerators coordinate with those who decided to submit an application to make sure that their application and payments are ready over the phone. They will agree on a date on which the application can be picked up. Once they are ready, an enumerator will contact the power house to remind them what you are doing and when they will be collecting the applications. [Note: In some cases, an employee of the power house will accompany the enumerator to help collect the applications and payments.]

On agreed date, an enumerator will visit habitation to collect application materials and go to power house. They will note for each households whether or not application was successfully submitted.

After 1 month and 3 months, enumerators will call respondents to observe electrification rates, and after 3 months they have household heads fill short endline questionnaire.

3.3 Data Collection

In order to collect our data, we will have experienced survey enumerators from Morsel India, an Uttar Pradesh-based survey company, to administer the surveys, distribute the information on the connection process, and assist in getting the household a connection. The surveys are administered using an Android smart phone app and are in Hindi. Baseline survey and immediate post-treatment questions about satisfaction with the campaign are conducted in person. Endline surveys are conducted over the phone, but will be followed up in-person as necessary.

3.4 Randomization

A list of 200 habitations for the study (additional 10 as back-ups). We will randomize treatment and control status by habitation on a spreadsheet, assigning pseudo-random numbers to each habitation. We then conduct block randomization by the electricity feeder for the habitation. This is done in order to ensure that quality of supply is balanced across treatment and control. The data is then sorted by the feeder name and the random number, and then the first half of habitations in each
feeder are assigned as treatment cases, with the remaining half assigned to the control.\textsuperscript{2} Habitations are chosen to match the specification laid out above. From each habitation, 10 households are selected to be surveyed in accordance with the above criterion.

\textsuperscript{2}In a few cases, our partners identified an odd number of habitations in a feeder. In these cases, the excess case is set aside as a spare to be used to replace a missing treatment or control habitation if it cannot be surveyed.
4 Outcome Variables and Estimation

4.1 Primary Outcome Variables

The primary outcome variables are binary indicators (yes/no) for whether the household fills out an application and whether they have a legal electricity connection at home. Given that households are non-electrified in the baseline, the outcome is only measured in the endline. The endline data for the indicator is collected through a mobile phone survey where possible and direct observation where the applicant submits an application through the campaign process.

We capture both the submission of an application and whether the connection was actually made, because there may be some issues that result in a divergence between these two outcomes, including: the electricity company fails to send someone to make the connection, the electricity company does not have proper supplies in stock (e.g. meters), or the respondent encounters some financial situation which makes it no longer viable for them to pursue a connection.

In our pilot surveys with this program, we encountered two cases where the electricity company failed to follow-up on an application within the period between submission and the endline survey, and one case of financial distress due to illness.

Our hypotheses based on these two variables are as follows:

**Hypothesis 1** Individuals who are exposed to the electrification campaign will be more likely to *submit applications* for electricity connections than those not exposed.

**Hypothesis 2** Individuals who are exposed to the electrification campaign will be more likely to *receive electricity connections* than those not exposed.

4.2 Secondary Outcome Variables

The electrification campaigns are also likely to change perceptions of the application process. Based on our previous work, we hypothesize that individuals overestimate both the cost and the difficulty of the electricity application process. If this is the case, we would expect those who are exposed to the campaigns to rate these difficulties significantly lower than those who are not part of the campaign.

**Hypothesis 3** Exposure to an electrification campaign decrease the perceived difficulty of
application for electricity access.

**Hypothesis 4** Exposure to an electrification campaign will decrease perceived cost of application for electricity access.

### 4.3 Understanding Causal Mechanisms

To understand the underlying causal mechanism, we look at several items. First, we directly ask those who receive an electricity connection how important the campaign was in gaining electricity access. This gives us a direct measure of the (perceived) importance of the campaign.

Within the campaign itself, we can also test what attributes of the campaign make it more likely for households to agree to submit an application. Households can agree to submit an application within two weeks, suggest that they do not want to submit an application now but are interested in doing so later, or can say that they never plan to submit an application.

We measure several characteristics related to the tutorial: their overall satisfaction with the tutorial, their rating of the difficulty of gaining electricity access, their rating of the cost of the application process, how much they learned in the tutorial, the extent to which they felt their questions were answered, and their perception of the household’s economic situation.

Agreeing to submit an application within two weeks can be modeled as

\[
Y_{ij} = \alpha + \beta X_{ij} + \zeta Z_k + \epsilon_{ij},
\]

where \(Y_{ij}\) is the indicator for whether household \(i\) in habitation \(j\) wants to submit an application in the next two weeks and \(\alpha\) overall intercept, \(X\) is a vector of the mechanism measures from the previous paragraph, and \(Z\) is a vector of fixed effects by feeder \((k)\). \(\epsilon_{ij}\) is the error term. We estimate robust standard errors clustered by habitation.

### 4.4 Model Specifications

Our first models deals with Hypotheses 1 through 4, dealing with the probability of applying for an receiving an electrical connection and the perceptions of obstacles. In both cases, a simple model...
for analysis can be specified as

\[ Y_{ij} = \alpha + \beta_1 \text{Campaign}_j + \gamma X_{ij} + \zeta Z_k + \epsilon_{ij}, \]  

(2)

where \( Y_{ij} \) is the indicator for whether household \( i \) submitted and application for and/or received a connection and \( \alpha \) overall intercept. Campaign is an indicator for whether the habitation \( j \) participated in an electrification campaign. \( X \) is a vector of control variables, including the participant’s (pre-campaign) estimate of electrification cost, their description of their current economic situation, whether they are aware of the “Saubhagya” plan for household electrification, and whether they have a ration card. \( Z \) is a vector of fixed effects by feeder \( k \). Finally, \( \epsilon_{ij} \) is the error term. We estimate robust standard errors clustered by habitation.

Finally, we also look for potential heterogeneous treatment effects based on the level of individual trust. We ask about trust in the state government, the utility company, the pradhan, and their neighbors. These are combined using a standard Likert Scale to produce an index of trust. Since high trust individuals are likely to be more willing to trust our enumerators to submit their applications and the government to follow through, we also run the following specification

\[ Y_{ij} = \alpha + \beta_1 \text{Campaign}_j + \beta_2 \text{Trust}_{ij} + \beta_3 (\text{Campaign}_j \ast \text{Trust}_{ij}) + \gamma X_{ij} + \zeta Z_k + \epsilon_{ij}, \]  

(3)

where \( Y_{ij} \) is the indicator for whether household \( i \) under habitation \( j \) submitted an application for and/or received a connection and \( \alpha \) overall intercept. Campaign is an indicator for whether habitation \( j \) participated in an electrification campaign. Trust is an indicator for high/low trust. \( X \) is a vector of control variables, including the participant’s (pre-campaign) estimate of electrification cost, their description of their current economic situation, whether they have a ration card. Finally, \( \epsilon_{ij} \) is the error term. We estimate robust standard errors clustered by habitation. The impact of the heterogeneous treatment effect will be evaluated using a standard F-test comparison of equation 3 to 2.
5 Power Analysis

In the summer of 2017, we conducted a small pilot of the electrification campaigns in three villages. We use this data to provide a realistic approximation of the effect of these campaigns on our response variables.

Because our sampling strategy involves binary individual decisions, nested in habitations, closed form power calculations can be difficult to develop. We therefore rely on simulation to determine the power of the design.

The algorithm creates 200 synthetic habitations. Submission of an application, receipt of electrification and naming of price as important obstacle are all modeled as a binomial process on the habitation level. The habitation level number of individuals who will experience positive outcomes is modeled as $n \sim b(x; n, p)$, and the habitation level probability of positive outcomes is $n/N$, where $N$ is the habitation population. For the algorithm, this means generating a count out of 100 from the binomial distribution and dividing by 100. Individual-level positive outcomes are then single draws from a binomial distribution with $p$ determined on the habitation level.

Once the individual and group potential outcomes have been assigned for both the treatment and control conditions, they are randomly sampled by habitation and the regression model is fit with clustered standard errors.

For the outcome of filling out an application, our field trials resulted in 6 of the 19 participants filling out an application, or $p = 0.32$. Given our observations about how many in the habitation had filled out an application, we do not anticipate that very many in the control condition will fill out an application. To be conservative however, we assume for the simulation that only one person fewer would have filled out an application, $p = 0.26$. This is a highly optimistic figure for the control condition. Still, as Figure 1 shows, we have 81% power to detect a significant relationship with our proposed sample size of 2,000.

In terms of the number of actual electrical connections, 3 of the 19 households in our pilot received connections, or $p = 0.16$. Again, we do not anticipate, based on our experience in the
pilot, that many people will receive connections without the intervention. We still estimate a relatively conservative test, assuming that only one fewer people would have received a connection without the intervention, $p = 0.11$. As Figure 2 shows, our sample size is still large enough to have detect a significant difference 87.8% of the time.

[Figure 2 about here.]

For trust in different outside actors, we did not notice substantial differences in trust in the government or the Pradhan. We did, however, see an increase in trust in the utility company, increasing from an average of 3.79 before the campaign to 4.27 at the endline (on a 1 to 5 scale). Even if we halve the size of the difference and assume a standard deviation of 1.06 (halfway between the sample standard deviations of the pre- and endline survey), we exceed 80% power with just 200 participants.
6 Study Timeline

- **December 28, 2017** Final list of candidate habitations compiled.

- **January 3, 2018** Randomization finalized.

- **January 5, 2018** Baseline begins and begin treatment administration.

- **January 31, 2018** Baseline surveys finished and all treatments administered.

- **February 10, 2018** Last of baseline data set to research team.

- **February 14, 2018** All applications from households collected.

- **March 14, 2018** First round of the endline survey completed.

- **March 24, 2018** Last of the endline data (first round) sent to the research team.

- **May 14, 2018** Second round of the endline survey finished.

- **May 24, 2018** Last of the endline data (2nd round) sent to research team.
7 Supplementary Material

We have attached to this pre-analysis plan the following supplementary materials:

- Questionnaires (PDF)
- Sample Habitations and Randomization (XLS)
- Cheat Sheet (For Habitations Covered by Saubaghy) (PDF)
References


Figure 1: Graph of power based on number of participants expected to fill out an application for an electrical connection, clustered within 200 habitations. It is assumed that the probability of uptake for the control condition is 0.32 and the probability for the treatment condition is 0.26.
Figure 2: Graph of power based on number of participants receiving an electricity connection, clustered within 200 habitations for receiving an electrical connection. It is assumed that the probability of uptake for the control condition is 0.11 and the probability for the treatment condition is 0.16.