Pre-analysis plan
"The Political Economy and Governance of Rural Electrification"[1]

AEA RCT Title: "The Political Economy and Governance of Rural Electrification in Kenya"

RCT ID: AEARCTR-0002389

Principal Investigators: Susanna Berkouwer, Eric Hsu, Edward Miguel, and Catherine Wolfram (University of California, Berkeley)

Date: February 28, 2019

Summary: The goal of this study is to quantify how three potential sources of accountability can improve governance in the provision of goods by the public sector, including one source of accountability that was experimentally varied by the researchers. This study employs both primary and secondary data sources. This document details the methodology used for the implementation of the experimental randomization, and outlines the intended analysis of a dataset containing information on construction quality at approximately 380 construction sites for low-voltage electricity infrastructure, as well as data on post-connection experience of approximately 1,500 households and firms located near these transformer sites. This document lays out the outcome variable definitions and main regression specifications that we intend to follow. We anticipate that we will carry out additional analyses beyond those included in this document, thus this document is therefore not meant to be comprehensive or to preclude additional analyses.

[1] This research has been supported financially by the Economic Development and Institutions (EDI) Initiative at the UK's Department for International Development (DFID) and the Development Impact Lab (DIL) at UC Berkeley.

Table of Contents:

Appendix:

- A. Monitoring letters
- B. Survey instruments
- C. Note on data management/access (uploaded on Oct 9, 2018)

1 Introduction

1.1 Summary

Governments in countries with low state capacity and poor institutions frequently outsource public goods provision to the private sector, but political economy and governance issues in this sphere remain poorly understood. This research will provide rigorous evidence on the private provision of public services in a low-capacity state. We study these issues in the context of Kenya’s national Last Mile Connectivity Project (LMCP), which aims to provide universal household electricity access by 2020 using World Bank and African Development Bank funding. We complement rich administrative data on the contracting process with novel field engineering assessments to study three channels affecting leakage and construction quality: (1) experimental variation in independent monitoring of the contractors; (2) donor aid conditionality; and (3) politicians’ electoral incentives. In particular, we will assess the extent to which third-party verification helps reduce collusion by improving enforcement, and whether donor conditions can limit clientelism. This pre-analysis plan focuses on the experimental variation created in channel (1) and discusses analyses of channels (2) and (3) as well as interactions among all three channels.

To understand the efficacy of independent monitoring as a source of government accountability, we randomize an independent scheme of construction monitoring as follows. We randomly selected a subset of construction sites and informed contractors that we would conduct engineering assessments at these sites post-construction, emphasizing that the results would be shared with international donors as well as Kenya Power and that negative construction outcomes may therefore affect their competitiveness for future construction contracts both in Kenya and internationally. The letters explicitly noted four items that we would assess. In fact, we inspected an additional 14 items for construction quality and numerous other items related to household experience with electrification. We describe this process in more detail in the sections “Randomization” and “Implementation of Treatment” below.

Our analysis employs both primary and secondary data sources. The secondary data sources were collected from Kenya Power starting June 2016 and continue to be collected throughout this project as construction progresses and data are updated. We launched primary data collection in September 2018. Since construction is concurrent to our surveys, we plan to continue data collection intermittently until around June 2020, depending on the speed of construction and survey work. We describe the surveying methodology, sampling, and instruments in more detail below.

1.2 Experimental design and implementation

1.2.1 Secondary data

In 2016-2018 we obtained administrative data from Kenya Power consisting of attributes of existing electricity infrastructure throughout Kenya, as well as progress reports of ongoing construction projects. For each transformer, these data include:

- GPS coordinates, with mappings to the county, constituency, and ward location
- GIS map of the infrastructure
- Engineering measurements
- The number of Kenya Power customers already connected
- For ongoing construction projects:
    - The contractor responsible for construction
    - Weekly or monthly progress reports (frequency varies by contractor) in eight distinct categories, ranging from "pole erection in progress" to "metering complete"
    - GPS maps of the proposed construction design and design specifications (e.g., pole material)
    - Number and location of new connections

We define a transformer community as all structures located within 600 meters of a transformer, in line with Kenya Power's general use of the term. We will continue to collect these secondary data, especially those related to ongoing construction projects, for the duration of the project.

1.2.2 Randomization

We used the secondary data obtained from Kenya Power to select our study communities. We restricted ourselves to transformers located in five counties -- Kakamega, Kericho, Kisumu, Nandi, and Vihiga -- because political affiliations vary widely in these areas with respect to government versus opposition support. In addition, we limited our sample to sites where stringing (installation of low voltage lines between poles) had not yet been completed as of August 2017. This left us with a sample of 632 communities to include. Using a computer random number generator, we then randomly assigned 320 sites to treatment and 312 sites to control, stratifying assignment by constituency and funder. For the primary data collection, we focus on a random subset of 380 out of the 632 communities.

1.2.3 Implementation of treatment

We identified five independent private sector contractors that were each responsible for a specific subset of the 320 treatment sites. During separate in-person meetings, we hand-delivered letters to individuals at the project management level of each of these contractors, informing them that the quality of construction at these sites would be inspected by an independent, international team of engineering surveyors. We emphasized that the results to these surveys would be shared with the World Bank, the African Development Bank, and Kenya Power, and may influence their future contracting opportunities. Each letter included the specific list of treatment sites that would be inspected and that this contractor was responsible for. It also included a sample list of the aspects of construction that would be inspected, such as low voltage wiring and electricity reliability post construction. A copy of the monitoring letter is included in Appendix A.

1.2.4 Field implementation

In September 2018, due to limitations in data collection funding, we randomly selected 60% (namely, 380 out of 632 communities) of the treatment and control sites for surveying, stratifying on funder and constituency, as described in detail below. Construction at sites funded by the African Development Bank was well underway by this point, while sites that were funded by the World Bank were only at the beginning stages of construction. We therefore selected 190 sites that were funded by the African Development Bank for our initial wave of surveying in Fall 2018.

On September 13, 2018, we launched a survey to capture engineering and socioeconomic outcomes at LMCP sites. See Appendix B for the full survey instruments.

- **Sampling poles**. At each site we randomize the branches of the network directly connected to the transformer. This randomization determines the order in which the branches are surveyed. The survey team completes all branches in the order of randomization. In cases where there is insufficient time to survey all branches, the survey team surveys as many branches as possible. For each branch that is surveyed, we collect data on all poles in the network (including all lines attached to the poles), up to 700 meters from the transformer. In addition, we collect more detailed technical information on a random subset of the surveyed poles, consisting of either 20 poles or one sixth of all poles at each site, whichever is greater. The main analysis will use data only from completed lines--in other words, for the main analysis we will discard data from any lines for which we only have partial data.

  As an additional analysis, we may include secondary data (from Kenya Power) from branches that were not surveyed and estimate the probability of a pole being surveyed as a function of distance from the transformer, and then use these probabilities as weights in the analysis. Poles on lines that were surveyed would all have equal weight.
- **Sampling occupants of connected structures**. At each site we document all drop cables along branches that are surveyed. We select structures by (1) randomly selecting drop cables in the network to survey, (2) surveying the primary occupant of the structure to which each sampled drop cable is attached.
- **Sampling occupants of unconnected structures**. We will record all unconnected compounds located within 100 meters of a surveyed pole in the low-voltage network for a given transformer. Among these compounds, we will survey two randomly selected unconnected compounds. If there are multiple households or firms at a selected compound, we will randomly select the occupant to be surveyed.

1.3 Analysis and data examined to date

At the time of registering this pre-analysis plan, the research team has obtained the following data:

- 2017 election data at the polling station level[2]
- 2013 election data at the constituency and ward levels
- Progress of LMCP sites in 2017 and 2018
- Socioeconomic data from 2009 Census
- Primary survey data from 62 transformer sites and 441 structure occupants.[3] By the end of the surveying activities in 2020, we plan to have collected primary survey data from all 380 transformer communities.

As described in the document titled, "Note on data management/access and pre-analysis plan," which was uploaded to the AEA RCT Registry on October 9, 2018 and is included in Appendix C of this document, any summary statistics or maps of key variables that have been collected to date were processed by others and have been viewed for the sole purpose of improving data quality and planning field work. The authors were not provided any information related to treatment status, except for data related to attrition and field planning, where treatment status was required to ensure that attrition and surveying rates are balanced across treatment and control. For a subset of the outcome variables, the authors have viewed summary statistics based on data collected from only the 36 control sites visited thus far. Each member of the research team agreed to follow the data management/access plan.

The remainder of this pre-analysis plan is organized as follows. Section 2 describes the main regression specifications, heterogeneity analysis, and planned methods of multiple hypothesis correction, in addition to other topics. Section 3 describes the major outcomes of interest. This document captures our current thinking about analysis with these data, but we anticipate carrying out some additional analyses beyond those included in this plan. As such, this plan is not meant to be an exhaustive set of all analyses we plan on carrying out, but rather a core set of initial estimates that will hopefully inspire further analyses.

[2] We thank Walter Mebane for sharing these data

[3] The authors of this pre-analysis plan have not accessed survey data directly, as prescribed by the data management plan that they have agreed to follow, see Appendix C.

2 Analysis

2.1 General notes

Throughout this document, we refer to our subject population as occupants of structures. Structures can be either connected or unconnected to the electricity grid. Structures can be used for residential or business/commercial purposes. In our setting, structures are typically located in compounds that can contain households, firms, or both. Our subject population consists of households or firms that were the main occupants of the structure that was attached to the drop cable (for connected structures) or a randomly chosen occupant of a selected unconnected compound (for unconnected structures).

We focus the main analysis on results where the occupant is the unit of analysis. However, we will also do analyses with outcomes defined at the drop cable, compound, or power meter level. If these results are different we will report them and explore why they differ. It is possible for a compound to have multiple households, for a drop cable to also supply power to structures outside the main compound, or for multiple households or firms to use power from the same meter. Conversely, it is possible for a household or firm to occupy structures that have multiple drop cables or multiple meters. Finally, some of our engineering metrics are collected at the level of the pole.

Throughout the remainder of this document, we refer to these households and firms collectively as "occupants."

2.2 Empirical specification – occupant-level outcomes

We test whether the experimental treatment affected occupant-level outcomes as follows:

$$y_{ic} = \beta_0 + \beta_1 T_c + X_c' \Lambda + Z_{ic}' \Gamma + \epsilon_{ic} \qquad (1)$$

where y_{ic} represents the outcome of interest for occupant i in community c, and T_c represents the treatment status of community c, which equals one if the transformer in that community was on the list provided to the LMCP contractor responsible for that site. This specification will test whether the outcome is affected by the intensity of third-party monitoring.

Following Bruhn and McKenzie (2009), we include a vector of community-level characteristics, X_c which includes the variables used for stratification during randomization (namely, constituency and funder) among other variables. In addition, we include Z_{ic}, a vector of occupant-level characteristics that includes gender and age of the respondent. Section 2.4 describes the components of these covariate vectors. Error terms will be clustered at the community level.

The sample consists of occupants at drop cables (approximately 6 per site) and unconnected compounds (approximately 2 per site) that are randomly chosen to be surveyed from the pole-tagging exercise. We will apply weights to each observation that is the inverse of the probability that it was sampled. For most occupants at drop cables, the probability of being sampled equals:

$$N_{cables\ at\ occupant} * \left(\frac{N_{cables\ visited\ by\ enumerator}}{N_{cables\ on\ poles\ assigned\ to\ enumerator}}\right) * \left(\frac{N_{branches\ surveyed}}{N_{branches\ at\ site}}\right)$$

Here, $N_{cables\ at\ occupant}$ represents the number of cables that are connected to buildings that are primarily used by the occupant being surveyed. $N_{cables\ on\ poles\ assigned\ to\ enumerator}$ represents the total number of cables in the portion of the network that was assigned to the enumerator to be surveyed. $N_{cables\ visited\ by\ enumerator}$ represents the number of randomly chosen cables that were visited by the enumerator. $N_{cables\ visited\ by\ enumerator}$ represents the total number of network branches leading away from the transformer at the site. $N_{cables\ visited\ by\ enumerator}$ represents the number of network branches leading away from the transformer that were surveyed at the site.

2.3 Empirical specification – community- and pole-level outcomes

To test whether the experimental treatment affected community-level outcomes (which are listed in Section 3), we will estimate an equation that is similar in form to the one in Section 2.2:

$$y_c = \beta_0 + \beta_1 T_c + X_c' \Lambda + \epsilon_c \qquad (2)$$

where y_c represents the outcome of interest for community *c*, and T_c represents the treatment status of community *c*. This specification does not include occupant-level covariates. Each community will receive equal weight in the analysis. Section 3 describes how we will construct the community-level outcomes.

For pole-level outcomes, we will estimate an equation similar to the one specified in Section 2.2:

$$y_{jc} = \beta_0 + \beta_1 T_c + X_c' \Lambda + \epsilon_{ic} \quad (3)$$

where y_{jc} represents the outcome of interest for pole j in community c, and T_c represents the treatment status of community c, which equals one if the transformer in that community was on the list provided to the LMCP contractor responsible for that site. We will apply weights to each pole that is the inverse of the probability that it was sampled. For most pole-level outcomes this probability is:

$$\left(\frac{N_{branches\ surveyed}}{N_{branches\ at\ site}}\right)$$

For more detailed outcomes that were collected for only a fraction of the poles at each site, this probability is:

$$\frac{1}{S_j} * \left(\frac{N_{branches\ surveyed}}{N_{branches\ at\ site}}\right)$$

Here, S_j is the skip interval the enumerator used to sample pole j for detailed assessment. At each site, each enumerator was instructed to conduct one detailed assessment for every S poles, where the integer S is chosen such that the enumerator will conduct a detailed assessment for 10 poles or one sixth of the poles that are assigned to the enumerator, whichever is greater. Note that none of our primary outcome indices will be constructed at the pole level. However, as described in Section 3, we intend to examine each component of the main indices individually. For pole-level outcomes we will conduct this analysis at the pole level rather than first aggregating to the community level.

2.4 Covariate vectors X_c, Z_{ic}

In this section, we describe each of the sets of covariates that we plan to utilize in the analysis. The vector X_c (community-level controls) includes the following variables:

- Constituency: Binary variables for each constituency indicating whether community c is in that constituency
- Funder: Binary variable indicating whether the transformer maximization at community c was funded by the World Bank or by the African Development Bank

- Ward rate of electricity usage for lighting (from 2009 Census)
- Ward socioeconomic index (from 2009 Census): Index that is constructed using percentage of residents in the ward with high quality walls (cement, brick, or stone), percentage of residents with high quality roofs (concrete, iron, or tile), percentage of residents with high quality floor (cement or tile), percentage of residents who have completed primary education, percentage of residents who have completed secondary education. To construct the index we will first normalize each component variable to have mean zero and unit variance across all wards in our study area. We will then take the sum of the normalized component variables.
- Number of months passed between construction start date and survey date

The vector Z_{ic} (occupant-level controls) will include the occupant-level variables listed below.

- Gender of respondent: Binary variable indicating whether the respondent is female.
- Age: Age of respondent at time of surveying.

2.5 Other sources of accountability

We hypothesize that there may be systematic differences in our results along additional differences between communities that were not randomly assigned. While it is difficult to identify causality in these cases, we will use statistical methods to study differences in our outcomes along these dimensions.

1. <u>Donor assignment:</u> 50% of sites in our sample are funded by the World Bank and 50% of sites are funded by the AfDB. During our conversations with partners at these organizations, as well as Kenya Power employees, we have found that there are significant differences in the way these donor organizations fund projects. The AfDB generally provides “turn-key” funding, meaning they provide a single contractor with all the funds needed to complete the designs, materials procurement, and construction work for all the sites that they are responsible for. The World Bank generally has stricter guidelines, requiring separate bidding processes so that different contractors may be responsible for these three components of the construction process. While the purpose of

these rigorous guidelines is to improve construction quality and minimize opportunities for corruption, donor community views are mixed about the effectiveness in achieving these goals, as well as the consequences of these requirements for timeliness and cost of construction. Based on multiple conversations with senior Kenya Power staff, the process of assignment of sites to each donor appears to be largely arbitrary, allowing us to credibly estimate differences in construction and socioeconomic outcomes between the two donors. To do this, in the regressions specifications above we will replace the treatment indicator, T_c, with an alternative treatment indicator, D_c, that equals one if the transformer maximization was funded by the World Bank.

2. Political alignment: our study area contains communities with a wide range of political orientations -- for example, results from the 2017 nationwide elections in Kenya show some constituencies in our study area with a more than 95% vote share for the government party and other constituencies reporting a more than 95% vote share for the opposition party. Other areas are considerably more mixed. Our main test will examine whether construction and socioeconomic outcomes differ between pro-government and pro-opposition areas. To implement this, in the regression specifications above we will replace the treatment indicator T_c with an alternative treatment indicator G_c that equals one if the transformer community is located in a government-supporting area. To determine political affiliation, we will use the vote share for Uhuru Kenyatta (the incumbent) from the 2013 Kenyan elections, which predate LMCP. If this vote share in a given area is greater than 50%, then we will classify this area as government-supporting. If the vote share is less than 50% then we will classify it as opposition-supporting. We will attempt to match ward level voting data to each transformer community, and use ward voting data to define community political affiliation. For the statistical analysis, we will include the control variables described in section 2.4 (Covariate Vectors). We will include constituency fixed effects, if feasible; however, if there is a very high degree of correlation in political affiliation across neighboring wards (and in the extreme, if all wards in a constituency share a political affiliation), we may be unable to include constituency fixed effects.

We will also study "contested" areas, defined as areas where at least two parties received at least 30% of the votes, to see if outcomes differ in these areas relative to more politically uniform areas. Finally, we will also use secondary data on construction progress from before and after the election to study how election dynamics affect outcomes. It could be that construction pace increased prior to the election, in order to win support, or it could be that construction pace increased post-election in areas where the government party gained the majority of the votes in order to reward political allegiance. Given that some part of infrastructure spending in Kenya is funded through local Constituency Development Funds (CDFs), which are controlled by local politicians, we will also assess the effect of variation in election outcomes for county assembly representatives, who are elected at the ward level. While we will primarily focus on 2013 election data to determine political alignment, we will also examine 2017 election data, since there were some shifts in regional political support between the two elections; however, the use of 2017 election data as a measure of political alignment is complicated by the possibility that promises of future LMCP assistance affected vote shares.

We will also explore how experimental treatment and the two other sources of accountability described above interact with one another. Do donor conditionalities, electoral incentives, and more intensive third-party monitoring substitute for or complement one another? To examine this, our main test will use interactions between T_c , D_c , and G_c . For occupant-level outcomes we will estimate the following equations that test for each pair-wise interaction. For community and pole-level outcomes, we will estimate equations that are analogous to equations (2) and (3).

$$y_{ic} = \beta_0 + \beta_1 T_c + \beta_2 D_c + \beta_3 T_c D_c + X_c' \Lambda + Z_{ic}' \Gamma + \epsilon_{ic}$$

$$y_{ic} = \beta_0 + \beta_1 T_c + \beta_2 G_c + \beta_3 T_c G_c + X_c' \Lambda + Z_{ic}' \Gamma + \epsilon_{ic}$$

$$y_{ic} = \beta_0 + \beta_1 D_c + \beta_2 G_c + \beta_3 D_c G_c + X_c' \Lambda + Z_{ic}' \Gamma + \epsilon_{ic}$$

We will also estimate an equation with all interactions, although we expect that for this test our sample size may limit statistical power to detect effects.

$$y_{ic} = \beta_0 + \beta_1 T_c + \beta_2 D_c + \beta_3 G_c + \beta_4 T_c D_c + \beta_5 T_c G_c + \beta_6 D_c G_c + \beta_7 T_c D_c G_c + X_c' \Lambda + Z_{ic}' \Gamma + \epsilon_{ic}$$

Lastly, we will explore heterogeneity based on the gender of the household head or firm manager surveyed. Female-headed households may be more likely to be excluded from traditional patronage networks. If such networks provide access to faster or higher quality construction, this would cause female-headed households to receive connections on average later and of lower quality. In addition, women may be politically and socially marginalized, so the quality of service provision even after construction may be different. By comparing female-headed with male-headed households and firms we will be able to test whether such patronage networks and marginalization lead to lower quality access to public services.

2.6 Construction of indices

When necessary, we will first aggregate the components of an outcome index such that they are all measured at the same unit of analysis. For example, if some components are measured at the pole level and some are measured at the community level, then we will first aggregate the pole level outcomes to the community level. The way we do this for each outcome is described in Section 3. Note that if all components of a particular index are measured at the same level then this step is not necessary. Next, we will normalize each component variable to have mean zero and unit variance. We will then construct the index by summing each normalized component variable (the mean effects approach). The occupant outcome indices will all be defined at the occupant level. When doing so will help preserve the sample size, we will replace missing values in the component variables with the mean in our sample.

Note that we will exclude any variables with zero variance since these do not contribute any information to the analysis. If a pre-specified variable is missing in more than 30% of respondents who were asked the relevant questions, then we will exclude it from the index. We cannot anticipate why a particular variable will be missing so frequently, but in such events where it warrants exclusion, we shall explore these reasons in the analysis and report results using the full list of components in the appendix. Also, for community-level proportions that we compute (e.g., proportion of poles that have a certain quality issue), we will drop the variable if over 95% of communities have a value greater than 0.95 or less than 0.05. We may also drop additional components of indices if it is found that they do not have meaningful variation. This may be because the vast majority of sites have little or no variation, or the variation is of a magnitude that is not meaningful from an economic or engineering perspective for that particular outcome. At the time of writing we have collected data for only a fraction of the AfDB-funded sites, making it difficult for us to predict what variation may be present in the final dataset--in particular, we have no information about variation among World Bank funded sites. In the event that we drop a component for this reason, we will explain the reasons and report results in which they are not dropped.

Finally, we will report results for all individual outcomes used to create indices in the appendix. The analyses for the individual outcomes will be conducted at the unit of the individual outcome. For example, we will aggregate pole-level outcomes to the community level for inclusion in an index that includes community-level outcomes. However, when analyzing the pole-level outcomes individually, we will do this analysis at the pole level (as described in Section 2.3). We will interpret results for individual components in light of the fact that the letters sent to contractors highlighted three specific construction outcomes that the Enhanced Monitoring Program would focus on: distance between poles, line sag, and quality of connection between transformer and LV wiring.[4] These outcomes correspond to: pole distance (1.1), line clearance from ground (1.5), bypassed fuses (1.2), and extended community-wide outages (1.18).

2.7 Multiple testing adjustment

In Section 3, we describe how the major outcomes of interest are categorized into eight primary outcome families and three additional secondary outcome families. For the main coefficient of interest (β_1), we will present two sets of p-values. First, we will present the standard "per-comparison", or naïve, p-value, which is appropriate for a researcher with an a priori interest in a specific outcome.

Second, since we test multiple hypotheses, it is also appropriate to control for the possibility that some true null hypotheses will be falsely rejected. Therefore, we will also present the false discovery rate (FDR)-adjusted q-value that limits the expected proportion of rejections within a hypothesis that are Type I errors (i.e., false positives). Thus, while a p-value is the unconditional probability of a Type I error, the analogous FDR q-value is the minimum proportion of false rejections within a family that one would need to tolerate in order to reject the

[4] A fourth outcome mentioned in the letters -- blackouts and electricity reliability post-connection -- is an important outcome in itself but does not correspond to any specific engineering measurements. Instead, blackouts and poor reliability can potentially be caused by multitude of reasons.

null hypothesis.[5] Specifically, we will follow the approach to FDR analysis adopted in Casey et al. (2012) and the references cited therein (e.g., Anderson 2008). We will present FDR-adjusted q-values for each of the eight outcomes within the primary outcomes group, as well as FDR-adjusted q-values for each component within each of the outcome families.

Section 3 below describes the primary outcomes and the outcome families that we will analyze. As noted in Section 1.4, we anticipate that we will examine additional outcomes beyond those included in this plan.

3 Major outcomes of interest

3.1 Overview

In this section, we specify eight primary outcomes of interest as well as secondary outcomes whose results we will also report. These outcomes have been selected based on the judgment of the research team and are arranged into several broad families, as described below.

Within each outcome family, there are outcomes at different levels of aggregation, ranging from specific variables to indices that combine data from multiple variables. Due to the novelty of many of these measures, some of the groupings are speculative. We will therefore report measures of index quality and coherence in the appendix, for example, by examining the correlation patterns of measures within each index. Depending on the index quality, we may also perform additional analyses, for example, presenting results with alternative groupings of outcomes. For completeness and transparency, in the appendix, we will also present estimated impacts for all specific outcomes individually, including those used to construct each of the indices.

3.2 Primary outcomes

Construction outcomes

1. Construction quality index

[5] In this sense, false positives are driven not only by sampling variation for a single variable (the traditional interpretation of a p-value) but also by having multiple outcomes to test.

2. Network size and configuration index
3. Construction timing index

Household outcomes

1. Household installation quality index
2. Household cost, experience, and bribery index
3. Reliability and safety index
4. Household knowledge index
5. Electricity usage index

3.3 Construction outcomes

We will analyze three indices representing: (1) network shape, (2) quality of construction, and (3) timing of construction. The network-shape index captures how different the constructed network is from the initial design. The construction quality index captures features of the construction that are below standard and are detrimental to safety or reliability of the grid. The construction timing index captures the amount of time passed before construction reached certain milestones.

For pole-level outcomes that are binary indicators, we will first aggregate to the community level by computing the proportion of poles with the indicated quality issue. For pole-level outcomes that are continuous measures, we will aggregate to the community level by taking the average over all poles for the site. As with all outcome indices, we will also report results based on the individual components (implementing the multiple hypothesis adjustment as described in section 2.7 above) and we will use these results to interpret findings based on the three main indices. For the construction quality index, we will re-sign each component as needed such that higher values represent higher quality.

An important outcome that is not listed below is the cost per connection for each transformer site. Although we have obtained cost data at the aggregated level for each contractor, we have not been able to obtain transformer-site level data on project costs. We will continue to try to obtain these data from Kenya Power and contractors. If possible, we will construct the best available estimates of total cost per connection at each transformer site, ideally disaggregated into material and labor costs. If we are able to construct cost as a transformer-level outcome, then we will include it as a primary outcome. Otherwise, we will discuss cost of construction at the most disaggregated level for which we are able to obtain data. If we are able to obtain costing data as well as contractor invoices at the transformer level, then we will also attempt to construct a measure of transformer project funds leakage, which would also be included as a primary outcome. (Currently, we are uncertain about the likelihood of obtaining the relevant cost data to construct these outcome measures at the transformer level.)

Outcome #1: Construction quality

Contractors may reduce costs by producing lower-quality construction, for example, by using cheaper materials, omitting construction elements, or using fewer resources to ensure proper installation. If we find that the quality index is higher at treated sites, we interpret this as evidence of cost cutting in the control group.

We will pay special attention to four quality components: pole distance (1.1), bypassed fuses (1.2), line clearance from ground (1.5), and extended community-wide outages (1.18). These four outcomes closely align with measures explicitly identified in the letter sent to contractors as elements that will be inspected. If positive effects tend to concentrate in these components but not in other markers of quality, this would suggest that treatment caused contractors to allocate resources to improve aspects of construction that they believed would receive extra external scrutiny. To test whether contractors diverted effort to improve the monitored construction outcomes, we will construct sub-indices for construction quality. One of these sub-indices will be constructed using only components 1.1, 1.2, 1.5, and 1.18. The other sub-index will be constructed using the remaining fourteen quality components. We will then test whether the treatment effects on these two quality indices differ from one another. We note, however, that many of the elements in the construction quality index could affect reliability, the fourth item explicitly mentioned in the letter, so a null result may not be conclusive in indicating that contractors diverted effort towards dimensions that they expected to receive greater scrutiny.

ID	Component	Unit	Description	Survey data used[6]	Contractor notified in letter
1.1	Pole distance	Pole	Distance between pole and previous pole in the direction of the transformer.	GPS coordinates (pole tagging survey)	X
1.2	Bypassed fuses	Transformer community	Binary variable that equals one if the transformer has at least one bypassed fuse	transformer_fuse_by (transformer survey)	X
1.3	Pole crack	Pole	Binary variable that equals one if the pole has a crack one centimeter or wider	pole_cracked	
1.4	Pole angle	Pole	Binary variable that equals one if the pole is leaning at an angle less than 85 degrees	pole_angle_yn, pole_angleestimate, pole_angle	
1.5	Line clearance from ground	Pole[7]	Binary variable that equals one if clearance between line and the ground is less than 6m at any point along the line	conductor_vert_yn, conductor_vert_m	X
1.6	Line clearance from objects	Pole	Binary variable that equals one if clearance between line and any objects is less than 0.5m	conductor_horiz_yn, conductor_horiz_m	
1.7	Lines cross	Pole	Binary variable that equals one if lines cross	conductor_lines_cross	
1.8	Pole cap missing	Pole	Binary variable that equals one if the pole is wooden and a pole cap is not present	pole_cap	
1.9	Pole circumference	Pole	Equals one if top of pole appears to be wider than the base	pole_diameter_yn, pole_diameter_1m, pole_diameter_2m	
1.10	Pole stability	Pole	Binary variable that equals 1 if pole is not stable	pole_stable	
1.11	Inadequate pole height	Pole	Binary variable that equals one if pole height is less than 8.5 meters	poleheight1	
1.12	Stay/strut	Pole	Binary variable that equals one if a stay or strut	stay_taut, stay_loc_correct,	

[6] All variable names refer to variables in the engineering survey unless otherwise noted

[7] The lines associated with each pole are the low-voltage lines leading to the previous pole in the direction of the transformer.

	improperly installed		was improperly installed	stay_angle_correct, strut_loc_correct, strut_angle_correct	
1.13	Stay/strut missing	Pole	Binary variable that equals one if a stay or strut was required at the pole but was not installed	stay_count, strut_count, stay_required	
1.14	Improperly installed insulator	Pole	Binary variable that equals one if an insulator is improperly installed	stay_insulator_distance_yn	
1.15	Missing insulator	Pole	Binary variable that equals one if an insulator was required but not installed	stay_insulated	
1.16	Service lines too long	Pole	Binary variable that equals one if a service line extends more than 15 meters from the pole	long_lines	
1.17	Percentage poles without grounding wire	Transformer community	Percentage of poles at the site that do not have a grounding wire	pole_grounding	
1.18	Frequency of outages	Transformer community	Proportion of months in which the community experienced an outage lasting at least two weeks. We will examine the first six to twelve months after LMCP construction is completed or the first six to twelve months for which we have data, whichever is later.[8]	(Transformer outage phone calls)	X

[8] We will attempt to collect outage data at each site in the 12 months following the completion of LMCP construction. However, for some sites, especially where LMCP construction is completed in late 2019 or 2020, we may be unable to collect a full 12 months of outage data. At those sites, we will attempt to collect at least six months of outage data and compute the frequency of outages in that shorter six to twelve month time frame.

Outcome #2: Network size and configuration

Contractors may be able to increase profits in at least two ways that can affect network size and configuration. First, contractors can drop parts of the network altogether to save on material, labor costs and time. This would cause us to observe design elements that are missing from the actual construction. If we find differences in the network shape index, we will examine differences in the individual components of the index to interpret whether differences seem to be driven by cost-cutting concerns (e.g., missing poles, lower coverage) or not. If we find differences in the pole count (2.2) or drop cable count (2.3) we will examine whether this is generally due to lower versus higher numbers of poles or drop cables in the construction as compared to the initial designs. It is also possible that there are fewer poles or drop cables constructed because contractors solicited bribes and did not follow through on construction unless they were paid the bribe. In addition, we will examine evidence from the household survey to help interpret differences -- for example, using local unpaid manual labor would be consistent with cost-cutting concerns as a motivating factor.

Second, contractors may connect ineligible structures or exert less effort to follow the design for other reasons, for example, in exchange for a bribe or to curry local political favors. These actions would be consistent with differences in network configuration, even in absence of missing material or lower coverage. We will examine evidence from household surveys to help interpret any differences -- for example, if asking for bribes were prevalent, this would support the story in which bribes facilitate changes in network shape. Note that the first three components measure deviations associated with some combination of cost-cutting and bribe-taking, and the fourth may capture deviations inspired by bribe-taking. The fifth measure (2.5) is likely to be associated with bribe-taking on the part of the contractor; to assist with interpretation, we will also be able to assess this by examining the household's response regarding bribe payments.

ID	Component	Unit	Description	Survey data used[9]
2.1	Percentage of compounds not electrified	Transformer community	Percentage of compounds within 600 meters of the transformer and within 100 meters of a low voltage line that are not connected to the grid	nlines, nuconns (pole tagging survey)
2.2	Difference in pole count as compared to design	Transformer community	Absolute value of the difference between the number of poles at the site and the number of poles in the initial KPLC design. We will normalize this difference in the pole count by dividing by the number of poles in the KPLC design.†	npoles (pole tagging survey)
2.3	Difference in drop cable count as compared to design	Transformer community	Absolute value of the difference between the number of drop cables at the site and the number of drop cables in the initial KPLC design. We will normalize this difference in the drop cable count by dividing by the number of drop cables in the KPLC design.†	nlines (pole tagging survey)
2.4	Construction-design graph edit distance	Transformer community	Number of steps needed to get from graph of initial KPLC design to graph of as-built construction. Inserting and deleting nodes costs one (deleting nodes from the design, however, is free), inserting and deleting edges costs one (deleting edges from the design, however, is free), relabeling nodes is free, and the transformer cannot be inserted deleted or relabeled. We will normalize this edit distance by dividing by the number of poles in the initial KPLC design	pole_id, pole_connected_id (pole tagging survey)
2.5	Percentage of poles beyond 600 meters from transformer	Transformer community	Estimated number of poles in the transformer community that are located more than 600 meters from the transformer	pole_id

† At some sites, due to time constraints we will have only surveyed K randomly chosen branches out of a total of N branches, with K<N. For these sites, we will compute this measure of deviation from design using the K branches in the design that minimize the deviation.

[9] All variable names refer to variables in the engineering survey unless otherwise noted

Outcome #3: Construction timing index

ID	Component	Unit	Description	Survey data used[10]
3.1	Construction start date	Transformer community	Months from January 2015 to the estimated date when transformer maximization started. (January 2015 is chosen as an arbitrary starting date to normalize time elapsed.)	date1a, date1b (background assessment)
3.2	Pole erected completed	Transformer community	Months from January 2015 to the estimated date when pole erection was completed*	date2a, date2b (background assessment)
3.3	Stringing completed	Transformer community	Months from January 2015 to the estimated date when stringing was completed*	date3a, date3b (background assessment)
3.4	Metering completed	Transformer community	Months from January 2015 to the estimated date when metering was completed*	date4a, date4b (background assessment)

*If pole erection happened in multiple waves, we will use the date when the first wave of pole erect was completed. We will treat stringing and metering similarly.

[10] All variable names refer to variables in the engineering survey unless otherwise noted

3.4 Household and Firm Survey outcomes

We will construct indices according to the following groups of outcome variables: Household installation quality index; Household cost, experience, and bribery index; Reliability and safety index; Household knowledge index; and Electricity usage index. For each family, we will report the index as the primary outcome, and then report individual estimates for each variable, implementing the multiple hypothesis adjustment as described in section 2.7 above. Where a family consists of two or more sub-families, we will report the family and the sub-family in the primary outcomes tables, and report the individual outcomes as well, possibly in the appendix. For each outcome family, we will re-sign each component as needed such that higher values represent better outcomes (e.g., higher quality, greater knowledge, better reliability). For all household survey outcomes except for Outcome #10 (firm performance), we will focus our main analysis on the households in the compound that have the drop cable that was randomly chosen to be surveyed. We will also restrict the sample to households connected to the grid after April 1, 2016, when Kenya Power officially began LMCP construction. Whenever data allow, we will also examine results when all connected households or firms with relevant data are included in the sample (which may require imputing data for some missing components in the outcome family).

For the following outcomes, we will Winsorize by setting observations in the top 1% to the 99th percentile: spending on wiring (5.3), upfront cost of connection (5.4), knowledge about debt to Kenya Power (7.3), electricity expenditures (8.4), electricity consumption (8.5), payments to date for connection (5.8), knowledge about total connection cost (7.1), and kerosene expenditures (9.11). Although the following outcomes focus on connected occupants, differential experiences of unconnected households are also of interest. Whenever the survey data allow, we will also present results for the unconnected households for individual components.

Outcome #4 – Household installation quality

ID	Component	Unit	Description	Survey data used
4.1	Has electricity ever flowed through this connection (either to the building where the drop cable terminates or to the building that has the meter)?	Drop cable	Equals one if electricity has flowed to the sampled drop cable.	intro8_electrified, intro8_electrifiedalt
4.2	Does this household have a meter?	Household	Equals one if the household has at least one meter	meters5
4.3	Has the meter ever worked?	Drop cable	Equals one if the meter has ever worked. Equals zero if there is no meter.	date2c
4.4	Does this household have a readyboard installed?	Household	Equals one if household has a readyboard installed	wire1
4.5	Did the contractor install any unrequested meters in this household? (fewer unrequested meters is better)	Household	Number of unrequested meters in the household	meters5, meter1_request
4.6	Time between submitting paperwork and receiving the electricity meter (shorter time is better)	Drop cable	Weeks after submitting paperwork but before receiving the electricity meter. We set this to the maximum value in the data if the meter was never received.	date12
4.7	Time between receiving the electricity meter and start of receiving electricity (shorter time is better)	Drop cable	Weeks after receiving electricity meter but before receiving electricity. We set this to the maximum value in the data if electricity has never been received.	date13

Outcome #5 – Household cost, experience, and bribery

ID	Component	Unit	Description	Survey data used
5.1	Number of days given to fulfill paperwork requirements	Drop cable		date11
5.2	Were you required to complete your own structure's wiring prior to being given a connection?	Drop cable	Equals one if structure needed to be wired in order to be	require_connect

			connected	
5.3	How much money did you spend on wiring (sum of labor costs and materials such as cables)?	Household	In Ksh	wire2a, wire2c, wire2d
5.4	How much did you pay in total, up front, to receive your electricity connection?	Drop cable	Total amount paid upfront for connection (in Ksh)	pay1
5.5	Who connected your household to the electricity grid?	Drop cable	Equals one if Kenya Power or REA connected the household to the grid	no_lmcp
5.6	In your experience getting connected to the grid, was anyone in your household ever asked to pay a bribe?	Household	Equals one if household was asked to pay bribe	pay1a
5.7	Did anyone in this household have to do any manual labor to assist with the erection of poles?	Household	Equals one if a household member had to do manual labor and was not paid for this labor	pay4, pay3a
5.8	How much has your household paid in total so far through installments for your connection?	Drop cable	In Ksh	pay1_paid
5.9	Generally speaking, how satisfied are you with the process of having electricity installed in your house?	Respondent	Measured on a five-point scale ranging from "very unsatisfied" (1) to "very satisfied" (5)	wtp3
5.10	In the past month, during how many hours did you experience very Low Voltage, to the point where you had to adjust your regular usage of your appliances or machines? (This question will be added to the survey starting in March 2019.)	Household	Integer (hours)	volt1
5.11	Think of all the appliances that were burnt or broken at any point in the past year because of electricity fluctuations. How much money did you have to spend when this happened, either to repair the broken device or to buy a new one? (This question will be added to the survey starting in March 2019.)	Household	In Ksh	volt4

Outcome #6 - Reliability and safety

a. Reliability

ID	Component	Unit	Description	Survey data used
6.1	Over the past 7 days, has your electricity connection delivered power to your household?	Drop cable	Equals 1 if yes	reliability1
6.2	Do blackouts occur regularly? [Yes/No question]	Household	Equals 1 if yes	f10e
6.3	Over the past 7 days, did your household experience any blackouts?	Drop cable	Equals 1 if yes	f10c
6.4	Over the past 7 days, for roughly how many hours was the power not working?	Drop cable		f10d
6.5	In the last year, have you experienced a power blackout which lasted over 30 days?	Household	Equals 1 if yes	long_blackout

b. Safety

ID	Component	Unit	Description	Survey data used
6.6	Have there been any accidents within the past year related to electricity causing injuries to humans or animals in the village?	Respondent	Equals 1 if the respondent is aware of one or more such accidents in the past year.	reliability4
6.7	Have there been any accidents within the past year related to electricity causing damage to any buildings in the village?	Respondent	Equals 1 if respondent is aware of one or more such accidents in the past year.	reliability6

Outcome #7 – Knowledge

a. Construction

ID	Component	Unit	Description	Survey data used
7.1	Before your household got connected, how much in total was your household told you would have to pay for the connection?	Household	Equals one if the household correctly said KES 15,000. Equals zero if the household does not know the correct connection fee or were never	pay1aa_amt

			told there was a connection fee.	
7.2	How frequently were you told you had to pay for the connection?	Household	Equals one if they correctly say monthly	pay1aa_freq
7.3	How much does your household still owe Kenya Power for your connection?	Household	Equals one 1 if they correctly say KES 15,000 minus whatever they've already paid in installments	pay1_owe

b. Electricity Consumption[11]

ID	Component	Unit	Description	Survey data used
7.4	20th token costs at least as much as 1st token.	Household	Equals one if respondent knows that the 20th token purchased in a month costs at least as much as the first token purchased in a month.	knowledge3, knowledge4a
7.5	Cost of first token	Household	Equals one if the cost for the first token, according to the respondent, is between 0.5 and 2 times the correct answer.	knowledge3
7.6	Cost of 20th token	Household	Equals one if the cost for the 20th token, according to the respondent, is between 0.5 and 2 times the correct answer.	knowledge4a

Outcome #8 - Electricity Usage

ID	Component	Unit	Description	Survey data used
8.1	Is your main source of lighting electricity?	Household	Equals 1 if household's main source of lighting is electricity	otherenergy1a
8.2	Is your main source of cooking electricity?	Household	Equals 1 if household's main cooking fuel is electricity	otherenergy2a
8.3	Has your household topped up since being connected to electricity?	Household	Equals 1 if household has topped up	usage1

[11] If Kenya Power changes the tariff structure during the course of the study, we will modify these three questions (7.4, 7.5, 7.6) in order to capture relevant dimensions of knowledge under the new policy. To account for changes to the tariff schedule we will count an answer as correct if it was correct within three months of the survey date.

8.4	Electricity expenditures over the past month: a. For post-paid customers: "How much was the amount in Ksh of your household's last monthly electricity bill?" b. For pre-paid customers: "In the past month, how much money in total did your household spend on top ups?"	Household		f8a, usage3a, connectedhh_usage_pre, connectedhh_usage_post
8.5	Electricity consumption over the past month	Household	Imputed electricity consumption based on reported expenditures on electricity in the past month	usage3a, f8a
8.6	Number of hours of lighting used at night in the past week	Household		usage0, usage7, usage8
8.7	Number of hours of lighting used in the morning in the past week	Household		usage00, usage9, usage10
8.8	Number of appliances owned that operate using the grid	Household		f19c
8.9	Number of households in this compound that are connected to the electric grid	Compound		intro10_nconnected

3.4.1 Secondary outcomes

The following three families contain secondary outcomes, beyond the eight primary outcomes we just described above (and that will be included in the main multiple testing adjustments). These three families of outcomes below are also of interest, although we do not anticipate meaningful effects on these outcomes over the timeframe of the study.

Outcome #9 - Household socioeconomic outcomes

All components are measured are on a 5-point scale: Strongly decrease, Somewhat decrease, No change, Somewhat increase, Strongly increase.

a. Quality of Life

ID	Component	Unit	Description	Survey data used
9.1	Has your grid connection allowed you to pursue non-agricultural forms of employment, self-employment, or start your own business?	Household		usage11
9.2	Has your grid connection affected your earnings?	Household		usage12
9.3	Has your grid connection permitted you to change the hours you work each week?	Household		usage13
9.4	Has your grid connection affected the amount of food you consume weekly?	Household		usage14
9.5	Do you believe your health has changed as a result of your grid connection?	Household		usage15
9.6	Has your grid connection affected your children's educational outcomes?	Household		usage16
9.7	Has your grid connection affected your knowledge about national and international news and current events?	Household		usage17
9.8	Has your grid connection affected your security; i.e., changed your exposure to crime?	Household		usage18
9.9	Has your grid connection enabled you to change the amount you spend on kerosene?	Household		usage19
9.10	Has your grid connection changed the frequency with which you charge your phone?	Household		usage20
9.11	In the past 7 days, how much money in total did your household spend on kerosene?	Household	In Ksh	otherenergy3a

b. Quality of Housing

ID	Component	Unit	Description	Survey data used
9.12	Owns home [=1]	Household		intro28
9.13	Number of rooms in primary residence	Household		wealth1
9.14	Materials of floors	Household	Equals 1 if high quality (contains cement or tile)	housequality1
9.15	Materials of roof	Household	Equals 1 if high quality (contains concrete, iron, or tile)	housequality2
9.16	Materials of walls	Household	Equals 1 if high quality (contains cement, brick, or stone)	housequality3
9.17	How many structures are in this compound?	Compound		n_buildings
9.18	How many structures are in this compound that are connected to electricity?	Compound		n_buildings_electrified

Outcome #10 - Firm Performance

ID	Component	Unit	Description	Survey data used
10.1	Does this business use electricity?	Firm	Equals 1 if yes	business2
10.2	Does this business plan to buy any electrical equipment in the next year?	Firm	Equals 1 if yes	business5
109.3	What does this business use electricity for?	Firm	Equals 1 if the business uses electricity for purposes other than lighting and cell phone charging	business3, bus_appli2
10.4	How many electrical appliances does this business own that use electricity from the grid?	Firm	Number of electrical appliances owned by the business	bus_appli2
10.5	Materials of roof	Firm	Equals 1 if high quality (contains concrete, iron, or tile)	busquality2

10.6	Materials of walls	Firm	Equals 1 if high quality (contains cement, brick, or stone)	busquality3

Outcome #11 - Political and Social Beliefs

ID	Component	Unit	Description	Survey data used
11.1	Considers "Electricity in households" one of two most important policies of the government of Kenya (Note: question is phrased as follows: *Of the following 7 government policies, which two do you believe are most important for Kenya: a) Electricity in households, b) Education, c) Reduce corruption, d) Better hospitals, e) Better roads, f) Fighting terrorism, g) Support for agriculture*)	Respondent	Equals 1 if yes	politics17a, politics17b
11.2	Do you think that the government is doing a good job ensuring that electricity is provided in Kenya? [=1]	Respondent	Equals 1 if yes	politics10
11.3	Did you vote in the August 2017 election? [=1]	Respondent	Equals 1 if yes	politics11a

3.4.2 Willingness to Pay outcomes

In addition to the indices reported above, we collect the following three measures of willingness-to-pay, for both connected and unconnected occupants:

1. Would you be willing to pay ${TotAmt} Ksh for an electricity connection?
2. Imagine that you were offered an electricity connection at ${TotAmt} today, and you were given 6 weeks to complete the payment. Would you accept the offer?
3. Would you be willing to pay ${MonthAmt} Ksh each month for a period of three years, so that you will pay ${TotAmt} Ksh in total?

Where ${} are randomly assigned amounts ranging from KES 0 to KES 35,000. We will report the demand curves generated by each of these measures of willingness to pay, for unconnected and connected occupants separately. Relatedly, we will also examine evidence for any differences in desire to be connected under LMCP (as captured by the variable "unconnected2" in the household survey).