LMCP Impact Evaluation
Pre-Analysis Plan

Dana Kassem and Giulia Zane
April 2021

1. Context and Project Summary

The Last Mile Connectivity Project (LMCP) is a mass electrification program in Kenya launched in 2015 by the Kenyan government with the goal of universal access to electricity. It is a multi-donor program that is being rolled out in multiple stages. The first phase started in 2016 and the African Development Bank was a major funder. The project covered all 47 counties in the country and it principally targeted low income households and small businesses. Specifically, in its first phase the project aimed to maximize the use of around 5,000 existing transformers, by extending the low-voltage network to reach households (1.2 million people) in the vicinity of these transformers. The target number of beneficiaries to be connected was 284,200 households and a minimum of 30,000 businesses by 2020.

The program was designed to maximize the use of existing Kenya Power and Lighting Company (KPLC) infrastructure. Specifically, the program maximized service areas of under-utilized transformers to include households and businesses within 600 m (in cable distance) of those transformers. Suitable transformers were identified based on their used capacity and their location. Another design feature was a subsidized connection fee of KSH 15,000. The subsidized connection fee is presented as a loan that is repaid as part of the prepaid tokens in case they decide to purchase electricity (i.e. for each amount paid up to 50% goes towards purchasing electricity units and the rest goes towards repaying the connection fee). In order to get a KPLC meter, households need to fill in an application form providing some personal information such as their ID number so that their KPLC account can be created. Moreover, households that already have electrical wiring in their house, also need to provide a certification of the quality of the wiring, issued by an authorized engineer. Those that do not have electrical wiring are provided with a “ready board” which allows them to plug in a light bulb and small electrical devices. Program beneficiaries purchase electricity through prepaid tokens that can be obtained using a mobile money account, a bank account, or purchased from local shops.

The aim of this study is to evaluate the impact of the LMCP on the living conditions of the beneficiaries. In what follows, we present our pre-analysis plan for conducting the study.

2. Design

2.1. Data Sources

We will mainly use primary data collected through an in-person survey of LMCP beneficiaries and suitable control households (see below for sampling strategy). We will use data from KPLC on the location of their LMCP transformers and cables to identify which households are the beneficiaries (geocoded meter level data is not available for all transformers/counties) and neighboring households that did not benefit from the program. We also plan to use KPLC administrative data on meter-level electricity consumption. In the future, we might explore using secondary data such as satellite data, the census of Kenya, or other readily available datasets provided that we can match them geographically to the location of LMCP transformers in a satisfactory manner.

2.2. Summary of the Identification Strategy

In order to credibly estimate the causal impact of LMCP on the socio-economic conditions of the beneficiaries, we exploit a unique design feature of the program, namely, the 600 m eligibility threshold. To be eligible for the subsidy, households or businesses need to be located within 600 m (in cable distance) of an LMCP transformer. Our identification strategy is a spatial regression discontinuity design where we compare eligible households within the 600 m boundary to control households just outside the eligibility border. The assumption is that households on either side of the boundary are similar in every aspect, except their eligibility for the program. We describe how we will implement this strategy in section 3 below.

2.3. Sampling Strategy

Our target sample size is around 7,450 households in 180 transformer communities in 6-10 counties. Our sampling strategy needs to take into account criteria that allow us to conduct the spatial RDD and maximize statistical power. Our setting is somewhat unique in the sense that we have information on the running variable before data collection as we will obtain distribution network maps from KPLC indicating both project lines and pre-existing lines. Using this information, we will adopt a spatial based household sampling method using satellite data, in particular, the population distribution map of Kenya from the Centre for International Earth Science Information Network (CIESIN)-Colombia University in collaboration with the Facebook Connectivity Lab and Digital Globe. We plan to use the following protocol to select transformers and sample households for our study. For each LMCP transformer:

1. We draw a buffer of radius 1 km around the transformer. That will constitute the overall catchment area of a transformer, or a *transformer community*. From that 1 KM buffer, we exclude the below in order to have our *final sample*:
    - 1.1. A buffer of 100 m of the transformer to avoid sampling households located too close to a market centre or another location of interest as they might be systematically different.
    - 1.2. A buffer of 40 meters of pre-existing lines (identifies as lines that were updated in the system before the beginning of the LMCP program and/or lines that have codes preceding the earliest that of the earliest LMCP line) to exclude, to the extent possible, households that were connected before the program.
    - 1.3. A buffer of around other LMCP transformers (600 m) to make sure we do not select households that are also treated by other transformers or are the control households for other LMCP transformers in the sample. As a rule of thumb, and to provide flexibility based on the density in different counties, the exclusion radius for LMCP I transformers with overlapping control regions will

be double the maximum distance between a control household and the LMCP I cables (treatment lines).

2. The sampling area for treatment households includes a buffer of 40 m around the LMCP cables in the transformer community.
3. The sampling area for control households includes a buffer of 300 m around the LMCP cables in the transformer community, as long as it falls within the 800 m buffer from the transformer.
4. The pool of structures (from the CIESIN database) that lie within the treatment and control sampling areas will be ranked randomly. Then, going through the ranked list, these structures will be validated using satellite images to determine whether these are real households or not. Enumerators will check them one by one until a maximum of 31 treatment and 31 control valid households are identified, the first 21 of which will constitute our main sample, and the remaining 10 will be our back-up sample in case the field visit to that household is unsuccessful.

The number of observations and distances defining treatment and control areas were chosen for Kakamega county. In case these are too restrictive in other counties, we might adjust them such that we get a sufficient sample. We will however be consistent across counties such that the thresholds and distances will be comparable.

It is important to note that our selection protocol yields a random sample of treatment and control households within a transformer community, but not a random sample of transformers. As mentioned above we require that a transformer community includes enough control and treatment households after various steps of elimination, which might yield a selected sample of transformers that are for example in less densely-populated areas, or have less transformers around. We will explore transformer characteristics after data collection in order to take this issue into account while interpreting our results and making statements about external validity. However, since our identification strategy is within a transformer community, it is the random sampling of the households within a community that matters, which means that our results will be internally valid.

2.4. Data cleaning and variable creation

At the time of registration of this pre-analysis plan, data collection has started in one county (Kakamega county). The sampling and data collection for the remaining counties will be done after the registration of this plan. Note that we did not perform any data analysis before registering this plan. We have explored data for 93 households to test whether the sampling protocol was selecting households from the correct target control and treatment groups, to check for errors in the programming of the survey instrument, and to improve or shorten certain sections that were taking too long in the field.

Because we are interested in the overall impact of the program and we have multiple outcomes, we might resort to summary indices that aggregate information from various variables. This aggregation is useful because it improves statistical power to detect impacts consistently across particular outcomes of interest. We will normalize specific outcome variables to have a mean 0 and a variance of 1, and then add up all the components to construct the index. We will exclude outcomes and controls that have zero variation since they will not be informative. Moreover, we will exclude from the index any pre-specified variable that is missing more than 30% of possible values. The final list of variables included in the indices will be reported in the appendix ex post. Finally, we will winsorize all continuous outcomes at the 99% level at the top and bottom, within each county. Outcomes that are a combination of multiple outcomes are computed based on unwinsorized components, and winsorized at the end.

3. Main Analysis

3.1. Main estimation equation

In order to identify the impact of the program on beneficiaries, we exploit the spatial discontinuity in the eligibility for the LMCP, namely the sharp change in access to the subsidized connection at 600 meter (cable) distance from the eligible transformers. With this sharp discontinuity, while not completely arbitrary, we can assume locally that the location of a household on either side of the 600 meter cutoff is “as good as random”. The identification assumption is that households on the two sides of the cutoff are comparable in every aspect except for their eligibility for the LMCP program. In practice, KPLC connected all eligible households within the 600m eligibility threshold, and the connection fee was subsequently deducted from their prepaid accounts (STIMA loan) once - and if- they start using electricity. Therefore, our sample could include eligible households who are connected but are not using electricity. Households just outside the 600 meter cutoff include both households that would get connected and use electricity and those who would not use electricity, ensuring that these are good counterfactual for the average outcomes within the 600 meter catchment area.

Let D_{ic} be the distance from the household to the transformer (or equivalently, distance to the eligibility border), and $E_{ic} \equiv D_{ic} <= 600$ the eligibility or RD indicator equal to 1 when a household lies within 600 m in cable distance of an LMCP transformer. Our main estimating equation at the *household level* is therefore:

$$y_{ic} = \alpha + \beta * E_{ic} + \theta * D_{ic} + \zeta * D_{ic} * E_{ic} + X'_{ic}\gamma + \delta_c + \epsilon_{ic} \qquad (1)$$

where y_{ic} is the outcome of interest of household i in community c, community FE δ_c, and covariates X_{ic} which we specify below. We will cluster standard errors at the transformer community level to account for potential correlation in outcomes for a community. At the individual level our estimating equation is:

$$y_{jic} = \alpha + \beta * E_{ic} + \theta * D_{ic} + \zeta * D_{ic} * E_{ic} + X'_{ic}\gamma + Z'_{jic}\rho + \delta_c + \epsilon_{ic} \qquad (2)$$

where Z'_{jic} is a vector of controls for individual $j$in household i community c.

3.2. Covariates/controls:

Community FE δ_c captures community-specific characteristics. In the vector of covariates X_{ic} , we will include household-level characteristics before the program. We don't have a baseline survey for our sample, but we will include in our survey questions regarding household characteristics in March 2016. These include:

ID	Variable	Type	Description	Unit	Ref.
1	Housing quality	Index	Index including :	HH	B27-

	index in March 2016		- Number of rooms - Indicator for high-quality floors - Indicator for high-quality roof - Indicator for high-quality walls - Distance to source of water		B34
2	Asset ownership index in March 2016	Index	Index including number of livestock (cattle, donkeys, sheep/goats, pigs) and means of transport (bicycle, motorcycle, car, tractor).	HH	B35-B35i
3	Indicator for living in the same house in March 2016	Indicator	Indicator for living in the same house in Mach 2016	HH	B26
4	Grid electricity in March 2016	Indicator	Indicator for being connected to KPLC electricity in March 2016	HH	C2-C2a
5	Solar system in March 2016	Indicator	Indicator for having a solar lighting system or solar home system in March 2016	HH	C68
6	Generator in March 2016	Indicator	Indicator for having a generator in Mach 2016	HH	C51
7	Rechargeable batteries in March 2016	Indicator	Indicator for having rechargeable batteries in Mach 2016	HH	C51
8	Number of appliances in March 2016	Count	Number of appliances owned in 2016 including: kitchen appliances, entertainment/IT appliances, and other appliances.	HH	C83, C85, C87

Individual level controls in Z'_{jic} will include gender, age, education (for individuals 18+), and indicators for being household head or respondent.

3.3. Running variable:

Our setting is unique in that we only observe the running variable (cable distance to the transformer) for treated households. By definition, we do not observe the cable line distance of the control households, because they are unconnected, and therefore there is no cable to measure. But since we know exactly who is treated and who is not, and what the threshold for the RDD exactly is, it is still possible to conduct the desired analysis in equation (1), by predicting the cable distance for control households. Based on discussions with KPLC engineers and LMCP contractors, the main predictors of cable length is the road distance and distance to pre-existing cables. We therefore plan to create a proxy for the cable distance of the control group by measuring the shortest road distance to the nearest constructed cable line (including the LMCP lines). In case suitable road data is not available, we will instead use the euclidean distance to the nearest constructed line. Let Δ_{ic} be the distance to the nearest line before connection (either road distance or euclidean if the former is not available). As before, D_{ic} is the actual cable distance for connected households to the transformer. We will estimate the following model with the sample of connected households:

$$D_{ic} = \alpha + \beta * \Delta_{ic} + \epsilon_{ic} \qquad (3),$$

to predict cable line distance for the control households $\widehat{D_{ic}} = \widehat{\alpha} + \widehat{\beta} * \Delta_{ic}$. Ideally we would like the correlation between D_{ic} and $\widehat{D_{ic}}$ to be high (or the R^2 of model (3) to be high). Based on interviews with KPLC and contractors, we expect this correlation to be high enough. However, if ex-post this is not the case, then we will try to improve the fit of model (2) by including potentially the road distance to the transformer, the euclidean distance to the transformer, and potentially higher order polynomials of these variables.

3.4. Heterogeneity Analysis

We plan to conduct secondary heterogeneity analysis that might shed light on the mechanisms through which electrification can affect socio-economic outcomes. We will do this by running the following regression:

$$y_{ic} = \alpha + \beta * E_{ic} + \theta * D_{ic} + \zeta * D_{ic} * E_{ic} + \eta_1 * E_{ic} * \chi_c \qquad (4)$$

$$+ \eta_2 * D_{ic} * \chi_c + \eta_3 * D_{ic} * E_{ic} * \chi_c + X'_{ic}\gamma + \delta_c + \epsilon_{ic}$$

where the characteristic of interest in represented by χ_c and is an indicator variable of whether that transformer community is above median in that heterogeneity dimension. The coefficient of interest would be η_1.

We specify below a short list of community-level characteristics that we would like to use in this heterogeneity analysis. It is possible however that the actual analysis would guide us into certain directions that are not pre-specified here.

Connection date is potentially an important factor in determining the impact of electrification on socioeconomic outcomes. It is possible that the longer the household has been connected to electricity, the higher the benefit they enjoy from electricity is. For example, it might take time for households to get familiar with how electricity can help them and hence for its benefits to accrue. In addition, households might need some time to save up and invest in electric appliances or other capital that is complementary to electricity. Therefore, we plan to explore heterogeneous outcomes based on how long the transformer community has been connected to electricity via LMCP.

Electricity usage is a second factor that could be important. In particular, since in our setting connection does not necessarily mean that the beneficiaries are using electricity, we could explore heterogeneous impact in transformer communities where electricity usage of the treatment is high versus in those where the usage is low.

Population density in the transformer community could be an important determinant of the impact of electrification. Location-based interaction between households depends on how close they live to each other, and hence, geographic spillovers could possibly be present in our setting. It is possible that an electricity connection affects not only the household itself, but also its neighbors and other households in the community. There are various categories of potential spillovers. The first category is *spillovers in electricity usage*: neighbors use electricity form a connected household. The second category is *spillovers in outcomes*. These can occur either through usage - neighbors can benefit from using the connected households electricity, (inviting people to watch TV, study, charge their phones, etc.), or through the market - by competing in labor or goods markets in which access to electricity would provide an advantage (increased access to information, possibility to operate in dark hours, etc.). Households very close to the border will also benefit from electrification, and their outcomes will be very close to the average outcome in the treatment area. The further away you go from the border, the smaller that benefit. Comparing average outcomes across the 600 meter discontinuity will lead to an underestimation of the true effect of electrification. To explore the presence of spillovers, we will conduct heterogeneity analysis based on population density in the transformer community. A potential data source for population density is the Gridded Population of the World dataset from CIESEN.

Finally, *distance to the county capital* is one dimension of heterogeneity that we are interested in as proximity to urban centers could be complementary to access to electricity.

3.5. Alternative Specifications

Depending on the availability of secondary data, we plan to conduct a spatial Difference in Differences (DID) type analysis using satellite data on luminosity, combined with the location of LMCP transformers in order to measure the effectiveness of the program. This rests on whether we are able to access data on all transformers in Kenya. We will conduct this analysis following the below steps:

1. Divide the counties in our sample into squares. In each year, classify each square into: (i) connected non-LMCP, (ii) connected LMCP, (iii) unconnected
2. The definitions above are based on the overlap of the squares created in step 1 with KPLC map data. For example, if a square contains an LMCP (non-LMCP) transformer, then it's classified as connected LMCP (connected non-LMCP). We can also do something more sophisticated: create a circle around the LMCP transformer, and for each square in Kenya we create an LMCP connectivity measure based on the share of the square overlapping with the circle. Same for non-LMCP transformers.
3. For each square created in step 1, have a measure of luminosity from the satellite data for every year
4. Run a FE panel regression at the square level with luminosity as the outcome and LMCP as treatment. This would be like a DID style specification:

$$L_{it} = \alpha + \beta * LMCP_{it} + \delta * NonLMCP_{it} + \theta_i + \gamma_t + \epsilon_{it} \qquad (5)$$

Similarly, we can potentially explore other left hand side variables to equation (5), in particular socio-economic outcomes, depending on data availability and the geographic resolution of these outcomes.

3.6. Multiple Hypothesis Testing:

Since we are interested in the impact of electrification on various socio-economic outcomes, we ought to account for Type I error (false positives) in hypothesis testing. False positives can occur because we have multiple outcomes and therefore multiple tests, each of which adding to the probability of false positives. We will present two p-values for each test, the standard p-value and the multiple inference adjust p-value. The first is a simple p-value without a correction for multiple hypothesis testing. This simple p-value would be relevant for readers interested in one particular outcome, for instance, the impact of electrification on self-employed income for women. Second, we will report the false discovery rate (FDR)-adjusted q-value, following Anderson (2008), Casey et. al. (2012), and Lee et. al. (2020).

4. Outcomes

We follow closely Lee et. al. (2020) in choosing our outcome variables. We start by listing primary outcomes that best represent the socio-economic conditions of interest in relation to the program and the particular context of Kenya. We then define 11 main categories of outcomes that we will focus on in our analysis.

4.1. Primary Outcome Variables

The list of outcomes below captures the main socio-economic conditions of interest.

ID	Outcome	Type	Description	Unit	Ref.
1	Any electricity used in past 30 days	Indicator	Used electricity in the past 30 days from KPLC, solar (excluding solar lanterns), generator or rechargeable batteries.	HH	C15, C32, C42, C55
2	Total spending on electricity from all sources	Total	Spending on KPLC electricity and other sources of electricity (generator, rechargeable batteries, solar) in the past 30 days	HH	C16, C39-C41, C49, C67, C69d
3	Employment	Proportion	Proportion of Employed or Self-employed (excluding agriculture) household members aged 18-70.	HH	D35, E5, E7
4	Monthly HH Income	Average income of hh (total HH	Salary for employed, revenue for self-employed, including agriculture	HH	D48, D48,

		income/no. adults)			E19, F13, F17, F22, F24
5	Female Employment	Proportion	Proportion of Employed or Self-employed (excluding agriculture) female household members aged 18-70.	HH	D35, E5, E7
6	Monthly Consumption	Average	Total household consumption in the past 30 days divided by no. of hh members	HH	J1-J 18
7	Respiratory health Index	Average	Average Index of overall symptoms experienced in the last 30 days (Persistent cough, Difficulty breathing, Asthma/ breathlessness at night, Eye irritation) by household members.	HH	D23-D25, D27
9	School enrollment	Indicator	For children under 18 not completed secondary school	Child	D2
10	Asset Ownership	Index	Index including number of livestock (cattle, donkeys, sheep/goats, pigs), means of transport (bicycle, motorcycle, car, tractor)	HH	B25-B25i
11	Life satisfaction	Index	Life satisfaction index (general life satisfaction, financial satisfaction and perceived safety)	Resp	H8-H 10
12	Political and current affairs awareness	Index	Political and current affairs awareness index based on 6 questions related to politics, education, and health	Resp	H1-H 6

4.2. Other Major Outcome Variables

4.2.1. Energy use

Was the LMCP successful in increasing access to electricity? The theory of how access to the electricity grid affects electricity usage is straightforward: without the LMCP program, households would have either to generate their own electricity (through the use of a small diesel generator or solar panels), pay the full connection fee if available to them from KPLC, receive it through other electrification programs (e.g. Kenya Off-Grid Solar Access Project (K-OSAP)), or finally live without electricity. Theoretically, the LMCP program should increase the beneficiaries' electricity use. In particular, we are not merely interested in whether the beneficiaries are connected to the grid or not, we are instead interested in whether actual electricity usage increased. This can be measured with variables such as ownership of electric appliances including for cooking and cooling/heating, lighting, fuel, etc., and night light satellite data. Households will be asked about other sources of electricity such as solar systems, which would allow us to estimate electricity use also for the households that are not connected to the electricity grid.

ID	Outcome	Type	Description	Unit	Ref.
1.1	Any electricity used in past 30 days	Indicator	Used electricity in the past 30 days from KPLC, solar (excluding solar lanterns), generator or rechargeable batteries.	HH	C15, C32, C42, C55
1.2	Total spending on electricity from all sources	Total	Spending on KPLC electricity and other sources of electricity (generator, rechargeable batteries, solar) in the past 30 days	HH	C16, C39-C41, C49, C67, C69d
1.3	Connected to KPLC	Indicator	Connected to KPLC grid with electricity	HH	C1
1.4	Use KPLC electricity for lighting?	Indicator	Indicator variable for whether the household uses KPLC electricity for lighting	HH	C19
1.5	Use Non KPLC electricity for lighting?	Indicator	Indicator variable for whether the household uses electricity from any source other than the KPLC grid (solar lantern, solar system, rechargeable batteries, generator) for lighting	HH	C33 C43 C53 C56
1.6	Use KPLC electricity for charging phones?	Indicator	Indicator variable for whether the household uses KPLC electricity for charging mobile phones	HH	C19
1.7	Use any electricity for charging phones?	Indicator	Indicator variable for whether the household uses electricity from any source other than the KPLC grid (solar lantern, solar system, rechargeable batteries, generator) for charging mobile phones	HH	C33 C43 C56
1.8	Total number of light bulbs	Total	Total number of light bulbs (grid and solar) used in the households	HH	C20.b C58
1.9	Share of light bulbs powered by KPLC grid	Proportion	Share of electric light bulbs powered by the grid (KPLC) out of the total number of light bulbs (KPLC + Solar)	HH	C20.b C58
1.10	KPLC electricity spending	Total	How much did the HH pay for electricity in the last 30 days?	HH	C16

1.11	Solar ownership	Indicator	1. Solar lantern 2. Solar lighting system 3. Solar home system/component based system	HH	C53
1.12	Total spending on energy	Total	Spending on KPLC electricity, other sources of electricity (generator, rechargeable batteries, solar) and other sources of energy (firewood, charcoal, kerosene, etc.)	HH	C16, C39-C41, C49, C67, C69d, C75, C80
1.13	Generator	Indicator	Did the household use a generator in the last 12 months?	HH	C29
1.14	Rechargeable batteries	Indicator	Did the household use a generator in the last 12 months?	HH	C29
1.15	Spending on other energy sources	Total	Spending in the last 30 days on: C40: Fuel of generator C49: Recharging batteries C67: Solar C69d: Dry-cell batteries C75: (*30/7) Firewood purchased Charcoal Kerosene Coal/lignite Animal waste/dung Crop residue/plant biomass/sawdust Briquette/pellet Biogas Ethanol	HH	C40 C49 C67 C69d C75

4.2.2. Income and employment

Access to electricity can increase the income and employment of beneficiaries for example by (i) providing them with new business opportunity that were not possible before (e.g. buy a mobile phone charging station or a fridge which they can exploit commercially), (ii) attract new businesses to the community which increases the supply of jobs, or (iii) getting better information about job availability through increased access to mobile phones or radio. Main outcomes will include wage income (agricultural and non-agricultural), self-employment income (agricultural and from commercial activities), employment status, and new employment, with particular focus on women’s employment.

ID	Outcome	Type	Description	Unit	Ref.
2.1	Employment	Proportion	Proportion of Employed or Self-employed (excluding agriculture) household members aged 18-70.	HH	D35, E5, E7
2.2	Monthly HH Income	Average income of hh (total HH income/no. adults)	Salary for employed, revenue for self-employed, including agriculture	HH	D48, D49, E19, F13, F17, F22, F24
2.3	Female Employment	Proportion	Proportion of Employed or Self-employed (excluding agriculture) female household members aged 18-70.	HH	D35, E5, E7
2.4	Employment - individual	Indicator	Indicator for whether the individual is Employed or Self-employed (excluding agriculture). For individuals aged 18-70.	Individ ual	D35, E5, E7
2.5	Employment-fema le	Indicator	Indicator for whether the individual is Employed or Self-employed (excluding agriculture). For women aged 18-70.	Woma n	D35, E5, E7
2.6	Total hours worked not agriculture - individual	Total	Number of hours worked for a wage + number of hours worked for a family business in past 7 days	Individ ual	D43, E8
2.6	Total hours worked not agriculture - woman	Total	Number of hours worked for a wage + number of hours worked for a family business in past 7 days	Individ ual	D43, E8
2.7	Has a business	Indicator	Indicator for whether the household owns a business	HH	E1
2.8	Business revenues	Total	Total revenues from household businesses	HH	E19
2.9	Business profits	Total	Total profits from household businesses	HH	E20
2.10	Employing workers	Indicator	Indicator for whether household businesses employ workers	HH	E9
2.11	Agricultural wages	Average	Total income obtained from agricultural wages divided by number of household members	HH	D48, D49
2.12	Non agricultural wages	Average	Total income obtained from non agricultural wages divided by number of household	HH	D48, D49

			members		
2.13	Agricultural revenues	Total	Total agricultural revenues	HH	F13, F17, F22, F24

4.2.3. Household Structure and Wealth

Household structure can be affected by electrification if for instance people's decision to migrate is affected.

ID	Outcome	Type	Description	Unit	Ref.
3.1	Household size	Total	Total number of household members	HH	B0
3.2	Female headed	Indicator	Indicator for female headed households	HH	A16 A18 A22
3.3	Migration or work away from village	Indicator	Whether at least one household member has migrated away from the village since March 2016 or at least one household member mostly slept away for work in the past 30 days.	HH	G21 B11
3.4	Housing quality	Index	Number of rooms Distance to water sources Indicator for high-quality floors Indicator for high-quality roof Indicator for high-quality walls	HH	B7 B9 B17 B18 B20
3.5	Asset Ownership	Index	Index including number of livestock (cattle, donkeys, sheep/goats, pigs) and means of transport (bicycle, motorcycle, car, tractor).	HH	B25-B25i
3.6	Savings increase	Indicator	Whether savings in the last 12 months have increased between March 2016 and today.	HH	G5
3.7	Borrowings increase	Indicator	Whether borrowings in the last 12 months have increased between March 2016 and today.	HH	G13

4.2.4. Health

Did the program reduce prevalence of injuries and respiratory illnesses? Did access to electricity affect fertility choices? Electrictric lighting and cooking reduces exposure to harmful indoor pollution from Kerosene lamps and cooking with firewood (Barron and Torero 2015). Outcomes will include health status of household members and whether injuries and illnesses prevented them from performing their regular duties, and fertility outcomes since the beginning of the program.

ID	Outcome	Type	Description	Unit	Ref.
4.1	Respiratory health Index (individual)	Index	Index of overall symptoms experienced in the last 30 days (Persistent cough, Difficulty breathing, Asthma/ breathlessness at night, Eye irritation) at individual level.	Individual	D23-D25, D27
4.2	Respiratory health Index (woman)	Index	Index of overall symptoms experienced in the last 30 days (Persistent cough, Difficulty breathing, Asthma/ breathlessness at night, Eye irritation) at individual level for female household members.	Woman	D23-D25, D27
4.3	Respiratory health Index (child)	Index	Index of overall symptoms experienced in the last 30 days (Persistent cough, Difficulty breathing, Asthma/ breathlessness at night, Eye irritation) at individual level for children <18	Child	D23-D25, D27
4.4	Experienced any illness or injury in past 30 days (individual)	Indicator	Indicator on whether the individual was ill or injured to the extent he/she was not able to perform his/her daily tasks for all household members.	Individual	D28
4.5	Experienced any illness or injury in past 30 days (woman)	Indicator	Indicator on whether the individual was ill or injured to the extent he/she was not able to perform his/her daily tasks for female household members.	Woman	D28
4.6	Experienced any illness or injury in past 30 days (Child)	Indicator	Indicator on whether the individual was ill or injured to the extent he/she was not able to perform his/her daily tasks for children <18	Child	D28
4.7	Fertility	Total	Number of children below 3 or current pregnancies in the HH	HH	D3 D30
4.8	Had malaria in past 30 days (individual)	Indicator	Indicator on whether the individual had malaria in the past 30 days for all household members.	Individual	D28

4.2.5. Consumption

An important measure of quality of life is how much households consume, especially in settings like ours where income measures are noisy.

ID	Outcome	Type	Description	Unit	Ref.
5.1	Monthly Consumption	Average	Total household consumption in the past 30 days divided by no. of hh members	HH	J1-J 18
5.2	Spending of food	Total	Spending on each of 17 food items in the last 7 days divided by no. of hh members	HH	J1
5.3	Monthly expenditure on non-durable goods and services	Average	Includes food, airtime, water, travel, hotels, clothing and shoes, recreation, personal items, cigarettes, gambling, contribution to savings account, religious expenses, medical expenses divided by no. of hh members	HH	J1-J 13
5.4	Annual expenditure on durable goods	Total	Includes household maintenance, entertainment equipment, household items (e.g. furniture, appliances, etc.), livestock, farming equipment divided	HH	J14-J18

4.2.6. Education

Theoretically, electricity can affect children's education through (i) longer study time because of better lighting, (iii) access to educational programs through the radio, and (iii) changing the incentives of parents to take the child out of school to work, both within household (e.g. no need to send child to collect firewood or water) and in the labor market.

ID	Outcome	Type	Description	Unit	Ref.
6.1	Study hours - total	Total	Total number of hours spent by children and young adults (age 7-22) in household studying during day and night hours in the last 3 week days	Child	D22
6.2	Study hours - total	Total	Total number of hours spent by children and young adults (age 7-22) in household studying during night hours in the last 3 week days	Child	D22
6.3	KCPE test score in 2018, 2019, and 2021	z-score	KCPE test score in 2019, and 2021 (if available)	Child	D15
6.4	KCSE test score	z-score	KCSE test score in 2019 and 2021 (if available)	Child	D12

6.5	School enrollment	Proportion	Proportion of children in the HH enrolled and attended in school	HH	D16, D16a

4.2.7. Productive use of electricity

Does electricity allow beneficiaries to engage in productive activities that are not feasible or are very costly without electricity? This category is designed to address such questions.

ID	Outcome	Type	Description	Unit	Ref.
7.1	Use electricity for business	Indicator	Indicator variable describing whether the household has at least one business connected to the KPLC grid	HH	E15
7.2	Household has a business that needs electricity to operate	Indicator	Indicator variable describing whether the household has a business that needs electricity to operate	HH	E16
7.3	Electricity used for agriculture.	Indicator	Indicator variable for whether electricity is used in irrigation, other agricultural activity, or the production or storage, of animal products or fish	HH	F7 F9
7.4	Number of electric appliances for the kitchen and other household chores	Count	Number of types of electric kitchen appliances (stove, oven, refrigerator, microwave oven, electric food processor (blender), rice cooker, freezer, kettle, and mill) and other electric appliances for household chores (electric iron, washing machine, and water pump (non-solar)).	HH	C82 C86
7.5	Number of heating, ventilation, and conditioning electric appliances	Count	Number of types of heating, ventilation, and electric conditioning appliances (fan, space heater, air condition, water heater)	HH	C86
7.6	Number of other electric appliances	Count	Number of types of electric appliances for lighting (security lights), IT and entertainment (radio, speaker/stereo, TV, satellite dish, computer/laptop, tablet, videogame, smartphone, brick phone, internet modem/router), and other (hair dryer, sewing machine)	HH	C84 C86

4.2.8. Time Use

the purpose of this category of outcomes is to understand whether electrification leads to more productive use of time (for example studying or producing goods for sale instead of spending more time on chores), with a particular focus on women and girls.

ID	Outcome	Type	Description	Unit	Ref.
8.1	Hours sleeping/resting	Hours	Sleep/Rest	Resp.	K2
8.2	Hours working	Hours	Household business Working for an agricultural wage Working for other sector wage	Resp.	K2
8.3	Hours farming	Hours	Farm work	Resp.	K2
8.4	Hours doing chores	Hours	Household chores Childcare Cooking Firewood Water	Resp.	K2
8.5	Hours doing chores by others	Hours	Average number of hours spent doing house chores by other members of the household or domestic helpers	HH	K4 K5
8.6	Hours enjoying leisure	Hours	Tv Radio Church Internet/Social media	Resp.	K2

4.2.9. Women empowerment

Did access to electricity improve women's bargaining power in the household? This can happen through economic empowerment of the woman in the household if electricity increases female labor supply (Dinkelman 2011).

ID	Outcome	Type	Description	Unit	Ref.
1.1	Woman working	Indicator	Indicator whether the woman has a job or a small business	Resp (W)	K7
1.2	Own savings	Indicator	Indicator whether the woman has her own savings in a financial institution	Resp (W)	K8
1.3	Household decision making	Index	Index of financial decision making in the household	Resp (W)	K9-K 12
1.4	School enrollment for girls	Proportion	Proportion of girls in the HH enrolled and attend in school	HH	D16, D16a
1.5	Woman is a decision maker for household business	Index	Index whether a woman in the household is a decision maker related to the household business (for households with businesses)	HH	E5, E6

4.2.10. Wellbeing and Knowledge

Subjective well being and happiness can be affected by access to electricity as people are more satisfied with their lives beyond economic and health improvements for example if they get utility from listening to the radio or having longer lighting. Electrification can facilitate information acquisition through access to media.

ID	Outcome	Type	Description	Unit	Ref.
10.1	Life satisfaction	Index	Life satisfaction index (general life satisfaction, financial satisfaction and perceived safety)	Resp	H8-H 10
10.2	Self-reported happiness	Indicator	Indicator equal to 1 if the respondent reports being happy or very happy	Resp	H7
10.3	General life satisfaction	Scale	Life satisfaction measured on a scale 1 to 10	Resp	H8
10.4	Financial satisfaction	Scale	Satisfaction about financial situation on a scale 1 to 10	Resp	H9

10.5	Perceived safety	Scale	Satisfaction about household safety on a scale 1 to 10	Resp	H10
10.6	Changes in life satisfaction compared to 2016	Index	Life satisfaction index in 2021 minus life satisfaction index in 2016	Resp	H7-H13
10.7	Political and current affairs awareness	Index	Knows the month and date the next general election Knows about the possibility of contracting COVID-19 in large gatherings not wearing a mask Knows whether it's possible to get COVID-19 by shaking hands Knows which students took the CBC curriculiculum between January and March 2021 Knows how many years secondary school will be with the new CBC system Knows who the current president of Tanzania is	Resp	H1-H6

4.2.11. Community-level variables

These community level variables will be used for descriptive analyses of the LMCP program and possibly for a heterogeneity analysis.

ID	Outcome	Type	Description	Unit	Ref.
11.1	Connection date	Median	Median date of connection in the treatment transformer community from KPLC's administrative data	Community	
11.2	Electricity quantity	Median	Median quantity of electricity consumed in the past three months treatment transformer community from KPLC's administrative data	Community	
11.3	Electricity Value	Median	Median value of electricity consumed in the past three months treatment transformer community from KPLC's administrative data	Community	
11.4	Travel time to county capital	Duration	Average time to travel from transformer to county capital	Community	
11.5	Population density	Density	Population density in the transformer community from CIESEN.	Community	

References

Anderson, Michael 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.

Casey, Katherine, Rachel Glennerster, Edward Miguel. 2012. “Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan.” *Quarterly Journal of Economics* 127(4): 1755-1812.

Dinkelman, T. (2011). The effects of rural electrification on employment: New evidence from South Africa. *American Economic Review* 101 (7), 3078–3108.

Lee, K., E. Miguel, and C. Wolfram (2020). Experimental evidence on the economics of rural electrification. *Journal of Political Economy* 128 (4), 1523-1565.

Olken, Benjamin A. 2015. "Promises and Perils of Pre-analysis Plans." *Journal of Economic Perspectives*, 29 (3): 61-80.