**Resilience to economic shocks through continued electricity access: evidence from Kenya**
**Pre-Analysis Plan**

Susanna B. Berkouwer, Eric Hsu, Oliver W. Kim, Kenneth Y. Lee, Edward Miguel, Catherine D. Wolfram
12 June, 2020

## Introduction

As COVID-19 spreads, access to electricity will be critical in allowing households and firms to continue productive activities, keep their phones charged, and stay up-to-date on the latest public health guidelines. At the same time, government restrictions needed to slow down the infection rate (e.g., stay-at-home orders, lockdowns, etc.) will cause severe economic impacts. We expect that poor communities will be severely affected, leading many to be unable to pay their bills. In Kenya, where millions of poor households were recently connected to the grid, we expect severe disruptions in people's access to affordable and reliable electricity.

A number of governments have recently announced electricity subsidy programs in response to the economic crisis, and more are considering implementation of such a program. To estimate the causal impacts of such subsidies during an economic crisis, we plan to implement a field experiment of electricity subsidies to poor households and firms in rural and urban areas of Kenya.

## Major research questions

- What is the impact of subsidized electricity on socio-economic outcomes?
- How does access to electricity affect resilience during an economic crisis?
- Do individuals cut back on electricity expenditures during an economic crisis?
- What is the willingness to pay (WTP) for electricity during an economic crisis?

## Sample

For this survey, we combine samples of previously surveyed individuals from two existing projects by this research team: the Rural Electric Power Project ("REPP"; see Lee, Miguel, Wolfram, 2020) and the Last Mile Connectivity Project ("LMCP"; ongoing work by Berkouwer, Hsu, Miguel, Wolfram, 2020). We define a site as all the buildings in a 700 meter radius of an electrical transformer. Sites are located in Busia, Kakamega, Kericho, Kisumu, Nandi, Siaya, and Vihiga counties, in the Western and Rift Valley regions of Kenya.

Our main experimental sample consists of 1,314 individuals spread across 268 sites who have pre-paid electrical meters. Due to movement restrictions related to COVID-19, the electricity subsidy treatments (detailed below) at the core of our experiment can only be administered to individuals who are (1) connected to the grid and (2) have a pre-paid electrical meter.

In addition, we will collect outcomes for a subset of 1,840 individuals from the combined REPP/LMCP samples who are either unconnected to the grid or do not have a prepaid meter. These individuals will allow us to answer additional research questions about the impact of COVID-19 and potential spillovers of electricity subsidies.

We are planning an additional, similar research activity concerning an urban sample of respondents based in Nairobi. We will provide further details on the urban sample in a separate addendum to this pre-analysis plan, to be pre-registered in the near future.

**Interventions**

This RCT will include one control group and three treatment arms designed to answer the research questions above.

T0. Control group

Respondents will be surveyed 3 times: once per month for 3 months. They will not receive any cash or pre-paid meter transfers.

T1. Treatment 1: Kenya Power transfer

Respondents will be surveyed 3 times: once per month for 3 months. In addition, the respondent will receive 500 Ksh (approx. 5 USD) in Kenya Power pre-paid electricity transferred to their meters once per month for 3 months.

T2A. Treatment 2A: Choice between Kenya Power subsidy or high cash transfer

Respondents will be surveyed 3 times: once per month for 3 months. In addition, the respondent will have a choice to either:

- Receive 500 Ksh (approx. 5 USD) in Kenya Power pre-paid electricity transferred to their meters once per month for 3 months.
- Receiving 500 Ksh (approx. 5 USD) in cash transferred to their mobile money accounts once per month for 3 months.

We will implement the respondent's choice.

T2B. Treatment 2B: Choice between Kenya Power subsidy or low cash transfer

Respondents will be surveyed 3 times: once per month for 3 months. In addition, the respondent will have a choice to either:

- Receive 500 Ksh (approx. 5 USD) in Kenya Power pre-paid electricity transferred to their meters once per month for 3 months.
- Receiving 350 Ksh (approx. 3.50 USD) in cash transferred to their mobile money accounts once per month for 3 months.

We will implement the respondent's choice.

If during initial surveying we find that more than 90% of respondents choose one of the options in T2A or T2B, we may adjust the cash amounts from 500 Ksh or 350 Ksh, respectively, to obtain a more informative set of responses.

**Assignment to treatment**

To minimize bias due to spillover effects, we will cluster at the site level for our sample. We assign the 268 sites where we have at least 1 connected respondent, to the treatments as follows:

- T0: 33% (90 sites)

- T1: 33% (89 sites)
- T2A: 17% (45 sites)
- T2B: 17% (44 sites)

We will stratify along two dimensions when assigning sites to treatment groups:

- Number of connected respondents at each site (ranging between 1 and 21)
- Whether the site was part of the REPP study (Lee, Miguel, Wolfram), the LMCP study (Berkouwer, Hsu, Miguel, Wolfram), or a newly enrolled "urban sample" participant based in Nairobi.

**Data Analysis**

We plan to conduct the following analyses. Note that the following is not an exhaustive list of the specifications: greater detail can be found in the pre-analysis plans for the related REPP and LMCP projects, linked below.

1. We will examine the impacts of the REPP experiment in which households at randomly chosen sites were incentivized to connect to electricity. The empirical specification will mirror those discussed in the pre-registered pre-analysis plan for the REPP project: https://www.socialscienceregistry.org/trials/350.
2. We will examine the impacts of the LMCP audit experiment, in which contractors were notified that randomly chosen sites would be monitored for construction quality. The empirical specification will mirror those discussed in the pre-registered pre-analysis plan for the LMCP project: https://www.socialscienceregistry.org/trials/2389.
3. We will examine the impacts of the electricity subsidy, described above. To do so, we will estimate the specification:
$$y_{si} = \beta_0 + \beta_1 T_{si} + X_{si}\Gamma + \varepsilon_{si}$$
where $y_{si}$ represents the outcome of interest for respondent $i$ at site $s$, $T_{si}$ equals 1 if the respondent is treated, and $X_{si}$ is a vector of control variables. The sample will include participants in the T0 and T1 arms. In case we find that over 90% of participants in the T2A or T2B arms of the experiment choose electricity tokens when using the initial cash amounts that are set, we will include those participants in this analysis as well.
4. We will examine how outcomes evolve over time to estimate the impacts of COVID-19.
5. We may also analyze the effects of different policies that the Kenyan government imposes in response to COVID-19. We will investigate the extent to which social, economic, and health behaviors and indicators react to government policies. The exact empirical approach will depend on the type of policy change and the scope of the area affected. For instance, if markets in one part of our study area are closed by the government, we will conduct a difference-in-difference analysis and compare changes in outcomes in areas where markets closed to changes in outcomes in areas without the policy.

We will winsorize all continuous outcomes at the 99% level at the top and bottom (or the top only, for outcomes where negative values are not logically possible). We conduct the winsorization within sampling groups (four arms of the REPP experiment, LMCP participants, and urban participants). Outcomes that are a combination of multiple outcomes are computed based on un-winsorized

components, and winsorized at the end. Whenever possible, we use baseline controls, which may come from the first round of surveys in May-June 2020. For some participants, we will be able to control for baseline values taken from previous surveys conducted for the REPP or LMCP projects. We also plan to conduct heterogeneity analyses. Across our major research questions, we will examine how these outcomes are shaped by other factors such as other sources of power (e.g. solar) and household wealth and income. For outcomes that are indices, for completeness we will also report effects on its individual components.

**Multiple Testing**

We will follow the approach to false discovery rate (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 outcome within the set of primary outcomes, as well as FDR-adjusted q-values for each component within each of the separate outcome families.

**Data Access**

As of this writing in May 2020, we have not merged in treatment status or estimated any impacts of treatment in the data. (Indeed, as of May 2020, we have not yet administered the treatment.) As described in the document titled, "Note on data management/access and pre-analysis plan," which was uploaded to the AEA RCT Registry in May 2020, the authors of this pre-analysis plan were provided with access to de-identified survey data that were stripped of any indicators that could expose the treatment status of households. We have used this access to make small improvements and fixes to the survey, and for a small set of hypothetical willingness-to-pay questions to set the cash transfer values in the T2A and T2B treatment arms.

**Primary outcomes**

We are pre-defining the following as **primary outcomes**. We regard these as particularly relevant for analysing treatment effects.

1. Electricity usage (EL2)
2. Total consumption expenditure (C4)
3. Total household income per capita (I6)
4. Total hours worked outside the home (self-emp & emp) (L4)
5. Total revenue for firms owned by household (E3)
6. Food security index (F1)
7. Child education index (CH4)
8. COVID Knowledge Index (CKN9)
9. COVID Symptoms Index (COV1)
10. Number of social interactions in last 2 weeks (NET5)
11. Mental Health Index (MH2)

**All Outcomes**

**A. Electricity**

We collect detailed information on household usage of electricity.

- EL1: Any electricity usage in the past week (=1)

- EL2: Continuous measure of electricity usage (= endline balance - baseline balance + EL5 + treatment)[1]
- EL3: Lost power because of unpaid bill in the past week (=1)
- EL4: Lost power because of unpaid bill in the past month (=1)
- EL5: Prepaid electricity expenditure since last survey round (only for survey rounds 2, 3)
- EL6: Prepaid electricity expenditure in the past 2 weeks
- EL7: Solar energy expenditure (solar home system + solar lanterns) in the past 2 weeks
- EL8: Other energy expenditure (kerosene + candles) in the past 2 weeks
- EL9: Number of electric appliances used for personal purposes in the past 2 weeks
- EL10: Number of electric appliances used for productive purposes in the past 2 weeks

## B. Consumption

We collect detailed information on consumption expenditure per capita at the household level[2]. Our primary outcomes are below:

- C1: Total food expenditure in the past 7 days: *4.1.1 + 4.3.Groceries/Food*
- C2: Total food consumption in the past 7 days: *C1 + 4.1.2a*
- C3: Non food expenditure in the past 7 days: *4.3.Household Items + 4.3.Assets + 4.3.Services + 4.3.Communication + 4.3.Housing + 4.3.Energy + 4.3.Transport + 4.3.Medical*
- C4: Total consumption expenditure in the past 7 days: *C1 + C3*

## C. Income

We calculate income per capita over the past 14 days, dividing household income by the number of members.

- I1: Agricultural earnings: 5.7
- I2: Self-employment earnings: sum of 5.23 across enterprises
- I3: Wage earnings: 5.27
- I4: Net Transfers: 5.33a - 5.34a + 5.35a + 5.36a + 5.37a
- I5: Dissaving: 5.32b + 5.34b - 4.3.Assets/Durables
- I6: Total household income: I1 + I2 + I3 + I4

## D. Labor supply and Wages

We calculate the following per household adults.

- L1: Agricultural labor supply. 5.2
- L2: Sum of 5.19 across all enterprises
- L3: Wage labor supply: 5.26
- L4: Total labor supply: L1 + L2 + L3
- L5: Child-care hours: 5.32

- W1: Average hourly wage: 5.27, 5.26 , average across all employed individuals, weighted by the number of hours worked in the last 14 days

[1] To ensure comparability across households surveyed at different times, we will normalize this by the number of days elapsed between baseline and endline rounds.

[2] This module was assigned to a random 50% subsample of households.

W1.1: Hourly wage in agriculture
W1.2: Hourly wage in non-agriculture

- W2: Self-employed hourly wage:
  W2.1: Self-employed hourly wage in agriculture
  W2.2: Self-employed hourly wage in non-agriculture

## E. Food security and Coping

- F1: Food security index: weighted index of answers to questions and subquestions 5.40-5.44 (negatively coded) and 5.45 (positively coded), calculated following Anderson (2008)
- F2: Food storage in kilograms: 5.4
- F3: Number of days household adults have skipped meals in the past 7 days: 5.43a
- F4: Number of days household children have skipped meals in the past 7 days: 5.43b

## F. Child education outcomes

- CH1: Any educational activity in the last 15 min: *3.12*
- CH2: Any educational activity in the last 24 hours: *3.13*
- CH3: Attended school in the last 24 hours: *3.13=attended school*
- CH4: Child education index (z-score of CH1, CH2, CH3)

## G. Enterprise outcomes / production

*Firm Operations / Revenue / Profits*

We analyse the following overall, as well as by sector of operation:

- E1: Net business creation: Net number of businesses operating within each bi-weekly and monthly period since January 2020 (5.16, 5.17c/d, 5.18a across enterprises).
- E3:[3] Agricultural production: 5.5 (after begin of the harvest period: 5.6 value) + 5.7
- E3: Total Revenue: 5.22
  Where an enterprise is no longer or temporarily not operating, we include these enterprises as having zero revenue.
- E4: Total Profits: 5.23
  Where an enterprise is no longer or temporarily not operating, we include these enterprises as having zero profits.

*Enterprise Labor*

- ENT1: Agricultural employment in the last 14 days: 5.3
- ENT2: Self-Employment in the last 7 days: 5.19 (extensive margin)
- ENT3: Wage employment: 5.20
- ENT4: Total employment: L1 + L2 + L3
- ENT5: Layoffs from the enterprises since January 2020: 5.21
  L5alt: Layoffs of household members since January 2020: 5.29
- ENT6: Unemployment rate in the past 14 days: 5.31
- ENT7: Agricultural hours: 5.2 (sum) + 5.3 * average agricultural hours in the past 7 days

[3] Labels occasionally skip numbers to remain consistent with the concurrent GE survey.

- ENT8: Self-employment hours in the past 7 days: 5.19 (sum)
- ENT10: Total Hours in the past 7 days: L7 + L8

## H. Social Trust, Violence and Networks

- SOC1: Most people can be trusted: Dummy if 9.1 = Most people can be trusted
- SOC2: Trusts government: Dummy if 9.2 = Somewhat trust or Strongly trust
- SOC3: Does not trust government: Dummy if 9.2 = Strongly distrust or Somewhat distrust
- SOC4: Crime indicator: Dummy if 9.3 = Yes
- SOC5: Crime types: Dummy for each of the categories in 9.3a
- SOC6: Outside visits in the past 14 days: 3.6
- SOC7: Outside travel in the past 14 days: 3.6a.Outsidevillage and 3.6a.Outsidesublocation
- SOC8: Outside visitors in the past 14 days: 3.7
- SOC9: Outside travel visitors in the past 14 days: 3.7a.Outsidevillage and 3.7a.Outsidesublocation

## I. Mental Health

- MH1: Life satisfaction (1-10 scale)
- CESD: sum of (non-refused) answers 6.6-6.15, correctly indexed so that negative outcomes correspond to higher values
- CESDa: For respondents with the shorter module:sum of the (non-refused) answers 6.1-6.5
- MH2: Combined index of CESD and CESDa, after standardized.

## J. Physical Health

- MOR: Mortality: Share of household members (by age group) that passed away each week / two weeks / month / since March 2019
- COV0: Indicator for covid-19 infection from self-reported symptoms: We use the existing public health literature to combine self-reported symptoms into a measure of covid-19 probability. If we are able to obtain longitudinal data from other contexts containing both self-reported symptoms and actual test results, we will use a machine-learning classification algorithm on that sample, and use the best-predictor on our self-reported symptoms.
- COV1: Sum of "yes" responses in 7.1
- COV2: Date of onset of symptoms: surveydate - 7.1a
- COV3: Current symptoms dummy: 7.1b
- COV4: Current symptoms dummy for other members: 7.2a
- COV6: Indicator if visited the hospital in the past 30 days. 7.4
- COV7: Number of visits to prenatal care. 7.7b.
- COV8: Water. Dummy if 7.9c reports negative change ("There is less water available", "I have to travel farther to obtain drinking water", "I have to pay more money for drinking water", or "The new source of water is dirtier")

## K. COVID-19 Knowledge and Behaviors Indicators

- CKN1: Has heard of virus: *8.1*
- CKN2: Symptoms counted: *sum of 8.3*
- CKN3: Correct prevention measures named: *sum of correct 8.4*
- CKN4: Incorrect prevention measures named: *sum of incorrect 8.4*

- CKN5: Has changed behavior: *8.5*
- CKN6: Correct behaviors: sum of recommended answers to *8.6*
- CKN7: Incorrect behaviors: sum of other answers to *8.6*
- CKN8: Feels nervous / anxious: *8.9*
- CKN9: Knowledge index: Weighted index of CKN3 (positively coded) and CKN4 (negatively coded)
- H1: Total medical expenditure in the last 7 days: *4.3.Medical Expenses OR 7.8/2*

## L. Network Indicators

We collect data on travel patterns and social contacts in our study area. For all households, we collect data on which places / cities / villages members of the household visited for market and / or religious activities. We also collect for everyone the total number of outside households somebody from the household visited and from which this household had visitors. We think of these as margin distributions capturing the extensive margin of social activity. In addition, we also collect data on the extent of in-person and not in-person social contacts.

We will look at the following primary outcomes over the past 14 days:

- NET1: Net migration of household members: *2.1 + 2.4*
- NET2: Attended religious activities: *3.4*
- NET3: Number of times visited market: *3.2*
- NET4: Number of households visited + number of visits received: *3.6 + 3.7*
- NET5: Number of interactions: *3.8*
- NET6: Number of offline interactions: *3.9*

In future work, we plan to use these network measures together with information on covid-19 symptoms to study the spread of the epidemic along network nodes.

**References**

Anderson, Michael L., ‘‘Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedaian, Perry Preschool, and Early Training Projects,’’ *Journal of the American Statistical Association*, 103, no. 484 (2008), 1481–1495.

Casey, Katherine, Rachel Glennerster, and Edward Miguel. "Reshaping institutions: Evidence on aid impacts using a preanalysis plan." *The Quarterly Journal of Economics* 127.4 (2012): 1755-1812.

Kenneth Lee, Edward Miguel, and Catherine Wolfram, "Experimental Evidence on the Economics of Rural Electrification," *Journal of Political Economy* 128, no. 4 (April 2020): 1523-1565.