Experimental Evidence on the Economics of Rural Electrification*

Kenneth Lee, University of Chicago
Edward Miguel, University of California, Berkeley and NBER
Catherine Wolfram, University of California, Berkeley and NBER

April 2019

ABSTRACT

We present results from an experiment that randomized the expansion of electric grid infrastructure in rural Kenya. Electricity distribution is a canonical example of a natural monopoly. Our experimental variation in the number of connections, combined with administrative cost data, reveals considerable scale economies, as hypothesized. Randomized price offers indicate that demand for connections falls sharply with price, and is far lower than anticipated by policymakers. Among newly connected households, average electricity consumption is very low, implying low consumer surplus. Moreover, we do not find meaningful medium-run impacts on economic and non-economic outcomes. We discuss implications for current efforts to increase rural electrification in Kenya, and highlight how credit constraints, bureaucratic red tape, low reliability, leakage, and other factors may affect interpretation of the results.

Acknowledgements: This research was supported by the Berkeley Energy and Climate Institute, the Blum Center for Developing Economies, the Center for Effective Global Action, the Development Impact Lab (USAID Cooperative Agreements AID-OAA-A-13-00002 and AIDOAA-A-12-00011, part of the USAID Higher Education Solutions Network), the International Growth Centre, the U.C. Center for Energy and Environmental Economics, the Weiss Family Program Fund for Research in Development Economics, the World Bank DIME i2i Fund, and an anonymous donor. We thank Francis Meyo, Susanna Berkouwer, Victor Bwire, Elisa Cascardi, Corinne Cooper, Stephen Harrell, Eric Hsu, Radhika Kannan, Anna Kasimatis, Tomas Monárrez, Emma Smith, Felipe Vial, and Catherine Wright for excellent research assistance, as well as colleagues at Innovations for Poverty Action Kenya and Remit Kenya. This research would not have been possible without our partners at the Rural Electrification Authority and Kenya Power. Hunt Allcott, David Atkin, Severin Borenstein, Raj Chetty, Carson Christiano, Maureen Cropper, Aluma Dembo, Esther Duflo, Garance Genicot, Sébastien Houde, Kelsey Jack, Marc Jeuland, Asim Khwaja, Mushfiq Mobarak, Samson Ondiek, Billy Pizer, Matthew Podolsky, Javier Rosa, Mark Rosenzweig, Manisha Shah, Jay Taneja, Duncan Thomas, Chris Timmins, Liam Wren-Lewis, and many seminar participants provided helpful comments. We are also grateful to the editor, Michael Greenstone, and four anonymous referees for extremely useful suggestions. All errors remain our own.

I. INTRODUCTION

Investments in infrastructure, including transportation, water and sanitation, telecommunications, and electricity systems, are primary targets for international development assistance. In 2018, for example, the World Bank directed a third of its global lending portfolio to infrastructure.[1] The basic economics of these investments—which tend to involve high fixed costs, relatively low marginal costs, and long investment horizons—can justify government investment, ownership, and subsequent regulation. While development economists have begun to measure the economic impacts of various types of infrastructure, including transportation (Faber 2014; Donaldson 2018), water and sanitation (Devoto et al. 2012; Patil et al. 2014), telecommunications (Jensen 2007; Aker 2010; Bjorkegren 2018), and electricity systems (Dinkelman 2011; Lipscomb, Mobarak, and Barham 2013; Burlig and Preonas 2016; Chakravorty, Emerick, and Ravago 2016; Barron and Torero 2017), there remains limited empirical evidence that links the demand-side and supply-side economics of infrastructure investments, in part due to methodological challenges. For instance, often it is not only difficult to identify exogenous sources of variation in the presence of infrastructure, but also to obtain relevant administrative cost data on infrastructure projects.

In this paper, we analyze the economics of rural electrification. We present experimental evidence on both the demand-side and supply-side of electrification, specifically, household connections to the electric grid. We compare demand and cost curves, and evaluate medium-run impacts on a range of economic, health, and educational outcomes to better understand the economics of mass rural electrification.

The study setting is 150 rural communities in Kenya, a country where grid coverage is rapidly expanding. In partnership with Kenya's Rural Electrification Authority (REA), we provided randomly selected clusters of households an opportunity to connect to the grid at subsidized prices. The intervention generated exogenous variation both in the price of a grid connection, and in the scale of each local construction project. As a result, we can estimate the demand curve for grid connections among households and, in a methodological innovation of the current study, the average and marginal cost curves associated with household grid connection

[1] In 2016 and 2017, the World Bank allocated over 40 percent of total lending towards its Energy and Extractives, Transportation, Information and Communications Technologies, and Water, Sanitation, and Waste Management sectors (World Bank Annual Report 2018).

projects of varying sizes. We then exploit the exogenous variation in grid connections induced by the randomized subsidy offers to estimate electrification impacts.

Household demand for grid connections is lower than predicted, even at high subsidy rates. For example, lowering the connection price by 57 percent (relative to the prevailing price) increases demand by less than 25 percentage points. The cost of supplying connections, however, is high, even at universal community coverage where the gains from economies of scale are attained. In our preferred specification using revealed preference data, estimated consumer surplus from grid connections is roughly one fifth of total construction costs. We derive a second measure of consumer surplus from a grid connection based on the subsequent benefits derived from consuming electricity, and this measure similarly implies low consumer surplus. In addition, we do not find economically meaningful or statistically significant impacts of electrification across a range of economic and non-economic (e.g., health, education, etc.) outcomes, collected in two rounds of surveys conducted roughly 16- and 32- months post-connection.

We next discuss several caveats in interpreting these results. First, the experiment generated a temporary reduction in the price of a grid connection. If credit-constrained households valued grid electricity services but were not able to raise the funds required to complete the purchase, the demand curve would underestimate the willingness to pay and thus consumer surplus. We present ancillary analyses from stated preference data on the potential importance of credit constraints in this context. We also consider the role of bureaucratic red tape and low grid reliability in reducing demand, and leakage in increasing construction costs.

Electricity systems serve as canonical examples of natural monopolies in microeconomics textbooks. Empirical estimates in the literature date back to Christensen and Greene (1976), who examine economies of scale in electricity generation. In recent decades, initiatives to restructure electricity markets around the world have been motivated by the view that while economies of scale are limited in generation, the transmission and distribution of electricity continue to exhibit standard characteristics of natural monopolies (Joskow 2000).

We differentiate between two separate components of electricity distribution. First, there is an access component, which consists of physically extending and connecting households to the grid, and is the subject of this paper. Second, there is a service component, which consists of the ongoing provision of electricity. There is some evidence of economies of scale in both areas. Engineering studies show how the costs of grid extension may vary depending on settlement patterns (Zvoleff et al. 2009) or can be reduced through the application of spatial electricity planning models (Parshall et al. 2009). With regards to electricity services, data from municipal utilities has been used to demonstrate increasing returns to scale in maintenance and billing (Yatchew 2000). Although recent work has examined the demand for rural electrification using both survey (Abdullah and Jeanty 2011) and experimental variation (Bernard and Torero 2015; Barron and Torero 2017), this is the first study to our knowledge that combines experimental estimates on the demand for and costs of grid connections, as well as the medium-run economic and non-economic impacts of grid connections. By combining these elements, we contribute to ongoing debates regarding the economics of rural electrification in low-income regions.

In Sub-Saharan Africa, roughly 600 million people currently live without electricity (IEA 2014), and achieving universal access to modern energy has become a primary goal for policymakers, non-governmental organizations, and international donors. In 2013, the U.S. launched a multi-billion-dollar aid initiative, *Power Africa*, with a goal of adding 60 million new connections in Africa. The United Nations Sustainable Development Goals include, "access to affordable, reliable, sustainable and modern energy for all." In Kenya, the government has recently invested heavily in expanding the electric grid to rural areas, and even though the rural household electrification rate remains relatively low, most households are now "under grid," or within connecting distance of a low-voltage line (Lee et al. 2016).[2] As a result, the "last-mile" grid connectivity we study has recently emerged as a political priority in Kenya.

At the macroeconomic level, there is a strong correlation between energy consumption and economic development, and it is widely agreed that a well-functioning energy sector is critical for sustained economic growth. There is less evidence, however, on how energy drives poverty reduction, and how investments in industrial energy access compare to the economic and social impacts of electrifying households. For rural communities, there are also active debates about whether increased energy access should be driven mainly by grid connections or via distributed solutions, such as solar lanterns and solar home systems (Lee, Miguel, and Wolfram 2016).

Although we find that the estimated consumer surplus from household grid connections is less than the total connection cost, universal access to electricity may still conceivably increase

[2] In 2014, the rural electrification rate in Kenya was 12.6 percent, according to the World Bank Databank (available at: http://data.worldbank.org).

social surplus.[3] For example, mass electrification may transform rural life in several ways: with electricity, individuals may be exposed to more media and information, might participate more actively in public life and generate improvements in the political system or public policy, and children could study more and be more likely to obtain work outside of rural subsistence agriculture later in life. However, roughly 16 and 32 months after being connected to the grid, rural Kenyan households show little evidence of any such gains, or their precursors. For instance, there are no meaningful impacts on objective political knowledge among respondents, nor on child test score performance. Of course, it is possible that the impacts of electrification take longer to materialize. Further long-run impact studies will thus be useful.

The remainder of this paper is organized as follows. Section II presents several natural monopoly scenarios that are empirically tested; Section III discusses rural electrification in Kenya; Section IV describes the experimental design; Section V presents the main empirical findings; Section VI offers an interpretation of these results, focusing on institutional and implementation challenges to rural electrification, and their implications; and the final section concludes.

II. Theoretical Framework

In the classic definition, an industry is a natural monopoly if the production of a particular good or service by a single firm minimizes cost (Viscusi, Vernon, Harrington 2005). More advanced treatments elaborate on the concept of subadditive costs, which extend the definition to multiproduct firms (Baumol 1977). Textbook treatments point out that real world examples involve physical distribution networks, and specifically cite water, telecommunications and electric power (Samuelson and Nordhaus 1998; Carlton and Perloff 2005; Mankiw 2011).

A. Standard model

We consider the case of an electric utility that provides communities of households with connections to the grid. To supply these connections, the utility incurs a fixed cost to build a low-voltage (LV) trunk network of poles and wires in each community. In the standard model, illustrated in figure 1, panel A, the electricity distribution utility is a natural monopoly facing high

[3] Note that we generally do not focus on "social welfare" because doing so would require imposing a particular social welfare function. Rather, we use the term "social surplus" throughout to capture the sum of consumer surplus from grid electrification, weighing all households equally, minus the costs of electrification.

fixed costs, constant or declining marginal costs, and a downward-sloping average total cost curve. As coverage increases, the marginal cost of connecting an additional household should decrease, as the distance to the network declines. At high coverage levels, the marginal cost is essentially the cost of a drop-down service cable that connects a household to the LV network. Household demand for a grid connection reflects expectations about the difference between the consumer surplus from electricity consumption and the price of monthly electricity service.

The social planner's solution is to set the connection price equal to the level where the demand curve intersects the marginal cost curve (p' in the figure). Due to the natural monopoly characteristics of the industry, the utility is unable to cover its costs at this price, and the social planner must subsidize the electric utility to make up the difference. In panel A, total consumer surplus from the electricity distribution system is positive at price p' since the area under the demand curve is greater than the total cost, represented by rectangle with height c' and width d'.

Note that we are assuming that, once connected, a household can purchase electricity at the social marginal cost. If this is true, there are no further social gains or losses from electricity consumption. An alternative approach to estimating the social surplus from a connection is to calculate the surplus from consuming electricity over the life of the connection. We implement this approach empirically in Section V.E.[4]

B. Alternative scenarios and potential externalities from grid connections

We illustrate an alternative scenario in figure 1, panel B. Here, the natural monopolist faces higher fixed costs. In this case, consumer surplus (the area underneath D) is less than total cost at all quantities, and a subsidized electrification program reduces social surplus.

In panel C, we maintain the same demand and cost curves as in panel B, but illustrate a case in which the social demand curve (D′) lies above observed private demand (D). There may be positive externalities (spillovers) from private grid connections, especially in communities with strong social ties, where connected households share the benefits of power with neighbors. In rural Kenya, for instance, people may spend some time in the homes of neighbors who have electricity, watching TV, charging mobile phones, and enjoying better quality lighting in the evening. Another factor that could contribute to a gap between D and D′ is the possibility that households have higher inter-temporal discount rates than policymakers. For example, if electrification allows children to

[4] Section I in Appendix A provides an additional discussion of the underlying theoretical framework.

study more and increases future earnings, there may be a gap if parents discount their children's future earnings more than the social planner. Further, observed private demand may be low due to market failures, such as credit constraints or a lack of information about long-run private benefits; what we call the social demand curve would also reflect the willingness to pay for grid connections if these issues were resolved. In general, if D′ lies above D, there may be a price at which the consumer surplus (the area underneath D′) exceeds total costs. In the scenario depicted in panel C, D′ is sufficiently high, and the ideal outcome is to offer full community coverage at price p''' and a subsidy equal to the rectangle with height $c''' - p'''$ and width d‴ provided to the utility.

Which of these cases best fits the data? In this paper, we trace out the natural monopoly cost curves using experimental variation in the connection price and in the scale of each local construction project, together with a combination of actual and estimated construction cost data provided by the electricity utility. The estimated cost curves correspond to the segments of figure 1 that range between the pre-existing rural household electrification rate level, which is roughly 5 percent at baseline in our data, and full community coverage (d=1). This is the policy relevant range for governments considering subsidized mass rural connection programs in communities where they have already installed distribution transformers.

One type of externality that we do not consider is the negative spillover from greater energy consumption, due to higher CO_2 emissions and other forms of environmental pollution. These would shift the total social cost curve up, making mass electrification less desirable. In the next section, we discuss aspects of electricity generation in Kenya that make these issues less of a concern in the study setting than they often are elsewhere.

III. Rural Electrification in Kenya

Kenya has a relatively "green" electricity grid, with most energy generated through hydropower and geothermal plants, and with fossil fuels representing just one third of total installed electricity generation capacity, which totaled 2,295 megawatts as of 2015. Installed capacity is projected to increase tenfold by the year 2031, with the proportion of electricity generated using fossil fuels remaining roughly the same over time.[5] Thus Kenya appears poised to substantially increase rural energy access by relying largely on non-fossil fuel energy sources.

In recent years, there has been a dramatic increase in the coverage of the electric grid. For instance, in 2003, a mere 285 public secondary schools (3 percent of the total) across the country had electricity connections, while by November 2012, Kenyan newspapers projected that 100 percent of the country's 8,436 secondary schools would soon be connected. The driving force behind this push was the creation of REA, a government agency established in 2007 to accelerate the pace of rural electrification. REA's strategy has been to prioritize the connection of three major types of rural public facilities, namely, market centers, secondary schools and health clinics. Under this approach, public facilities not only benefited from electricity but also served as community connection points, bringing previously off-grid homes and businesses within relatively close reach of the grid. In June 2014, REA announced that 89 percent of the country's 23,167 identified public facilities had been electrified. This expansion had come at a substantial cost to the government, at over \$100 million per year. The national household electrification rate, however, remained relatively low at 32 percent, with far lower rates in rural areas.[6] Given this grid expansion, the Ministry of Energy and Petroleum identified last-mile connections for "under grid" households as the most promising strategy to reach universal access to power.

During the decade leading up to the study period, any household in Kenya within 600 meters of an electric transformer could apply for an electricity connection at a fixed price of \$398 (35,000 KES).[7] The fixed price had initially been set in 2004 and was intended to cover the cost of building infrastructure in rural areas. As REA expanded grid coverage, the connection price emerged as a major public issue in 2012, appearing with regular frequency in national newspapers and policy discussions. The fixed price seemed out of reach for many if not most poor, rural households to afford (annual per capita income is below \$1,000 for most rural households).

---

[5] Specifically, in 2015, total installed capacity consisted primarily of hydro (36 percent), fossil fuels (35 percent), and geothermal (26 percent) sources. Based on government planning reports (referred to as *Vision 2030*), total installed capacity is expected to reach 21,620 MW by 2031, with fossil fuels (e.g., diesel and natural gas) representing 32 percent of the total. Many other African countries generate similar shares of electricity from non-fossil fuel sources (Lee, Miguel, and Wolfram 2016).

[6] REA provided us with estimates of the proportion of public facilities electrified (June 2014), the national electrification rate (June 2014), and overall REA investments (between 2012 and June 2015).

[7] All Kenya Shilling (KES) amounts are converted to U.S. dollars at the 2014 average exchange rate of 87.94 KES/USD. All 2016 and 2017 KES amounts are first adjusted to 2014 levels using the appropriate inflation rate before converting to USD. The fixed price of 35,000 KES was established in 2004 to reduce uncertainty surrounding cost-based pricing. Anecdotally, it was common for service providers to lower the cost-based price in exchange for a bribe.

However, Kenya Power, the national electricity utility, held firm, estimating the cost of supplying a single connection in a grid-covered area to be far higher at $1,435. After the government rejected its proposal to increase the price to $796 (70,000 KES) in April 2013, Kenya Power initially announced that it would no longer supply grid connections in rural areas at all, limiting supply to households that were a single service cable away from an LV line. As a result, the government agreed to temporarily provide Kenya Power with subsidies to cover any excess costs incurred, allowing the expansion of rural grid connections to continue at the same $398 price as before. In February 2014, the government ended these subsidies to Kenya Power, and it was again widely reported that the price would increase to $796. Ultimately, the $398 fixed price remained in place for households within 600 meters of a transformer throughout the first phase of the study period, from late-2013 to early-2015, when study subsidies for electric grid connections were distributed and redeemed.

The government announced in May 2015 (after baseline data collection activities and redemption of most subsidy offers) that it had secured $364 million—primarily from the African Development Bank and the World Bank—to launch the *Last Mile Connectivity Project* (LMCP), a subsidized mass electrification program that plans to eventually connect four million "under grid" households, and that, once launched, would lower the fixed connection price to $171 (15,000 KES). This new price was based on the Ministry of Energy and Petroleum's internal predictions for take-up in rural areas, and was revealed publicly in May 2015. The take-up data described in the next section were collected during the decade-long $398 price regime, and before any public announcement of the planned LMCP program.

IV. EXPERIMENTAL DESIGN AND DATA

A. Sample selection

The field experiment takes place in 150 "transformer communities" in Busia and Siaya, two counties that are typical of rural Kenya in terms of electrification and economic development, and where population density is fairly high (see appendix table B1). Each transformer community is defined as all households located within 600 meters of a secondary electricity distribution (low-voltage, LV) transformer, the official distance threshold that Kenya Power used for connecting buildings at the standard price. The communities were sampled in cooperation with REA.[8]

Between September and December 2013, teams of surveyors visited each of the 150 communities to conduct a census of the universe of households within 600 meters of the central transformer. This database, consisting of 12,001 unconnected households in total, served as the study sampling frame, and showed that 94.5 percent of households remained unconnected despite being "under grid" (Lee et al. 2016).

Although population density in this setting is fairly high, the average minimum distance between structures is 52.8 meters.[9] These distances make illegal connections quite costly, since local pole infrastructure would be required to "tap" into nearby lines; in practice, the number of illegal connections is negligible in the study sample (unlike in some urban areas in Kenya).

For each unconnected household, we calculated the shortest (straight-line) distance to an LV line, approximated by either the transformer or a connected structure. To limit construction costs, REA requested that we limit the sampling frame to the 84.9 percent of households located within 600 meters of a transformer that were also no more than 400 meters away from a low-voltage line.[10] Applying this threshold, we randomly selected 2,289 "under grid" households, or roughly 15 households per community.

B. Experimental design and implementation

Between February and August 2014, a baseline survey was administered to the 2,289 main study households. We additionally collected baseline data for 215 already-connected households, or 30.5 percent of the universe of households observed to be connected to the grid at the time of the census, sampling up to four connected households in each community, wherever possible.[11]

In April 2014, we randomly divided the sample of transformer communities into treatment and control groups of equal size, stratifying the randomization process to ensure balance across county, market status, and whether the transformer installation was funded early on (namely, between 2008 and 2010). The 75 treatment communities were then randomly assigned into one of

[8] See Section II in Appendix A for further details, and appendix figure B1 for a map of the sample communities.
[9] A map of a typical transformer community (in terms of residential density) illustrating the degree to which unconnected households are under grid is presented in appendix figure B2.
[10] In other words, all households located within 400 meters of the transformer were included in the sampling frame, while some households located between 400 to 600 meters of the transformer were excluded.
[11] A summary of the experimental design is provided in appendix figure B3.

three subsidy treatment arms of equal size. Following baseline survey activities in each community, between May and August 2014, each treatment household received an official letter from REA describing a time-limited opportunity to connect to the grid at a subsidized price.[12] Households were given eight weeks to accept the offer and deposit an amount equal to the effective connection price (i.e., full price less the subsidy amount) into REA's bank account.[13] The treatment and control groups are characterized as follows:

1. High subsidy arm: 380 unconnected households in 25 communities are offered a $398 (100 percent) subsidy, resulting in an effective price of $0.
2. Medium subsidy arm: 379 unconnected households in 25 communities are offered a $227 (57 percent) subsidy, resulting in an effective price of $171.
3. Low subsidy arm: 380 unconnected households in 25 communities are offered a $114 (29 percent) subsidy, resulting in an effective price of $284.
4. Control group: 1,150 unconnected households in 75 communities receive no subsidy and face the regular connection price of $398 throughout the study period.

Treatment households also received an opportunity to install a basic, certified household wiring solution (a "ready-board") in their homes at no additional cost. Each ready-board—valued at roughly $34 per unit—featured a single light bulb socket, two power outlets, and two miniature circuit breakers.[14] Each connected household was fitted with a prepaid electricity meter at no additional charge. At the end of the eight-week period, treatment households could once again connect to the grid at the standard connection price of $398.

After verifying payments, we provided REA with a list of households to be connected. This initiated a lengthy process to complete the design, contracting, construction, and metering of connections: the first household was metered in September 2014, the average connection time was

[12] An example of this letter is provided in appendix figure B4.
[13] Note that in this setting, one does not need a bank account to deposit funds into a specified bank account. The high subsidy (free treatment) group described below is not subject to the additional ordeal of traveling to town to access a bank branch, and interacting with bank staff to deposit funds into REA's account. For households that need to pay for a connection, the total time and transport cost of such a trip is roughly a few hundred KES (or a few U.S. dollars), far smaller than the experimental subsidy amounts.
[14] The ready-board was designed and produced for the project by Power Technics, an electronic supplies manufacturer in Nairobi. A diagram of the ready-board is presented in appendix figure B5.

seven months, and the final household was metered over a year later, in December 2015.[15] Additional details are discussed in Section VI.B.

Between May and November 2016, we administered a first follow-up survey ("R1") to 2,217 study households, or 96.9 percent of the baseline sample. We also surveyed an additional 1,328 households—between six to eleven households per community—as part of a "spillover sample," randomly sampling households that were observed to be unconnected at the time of the census but were not chosen for the baseline survey. Furthermore, we administered short language and math tests to all 12 to 15-year old's in the sample, or 2,302 children in total.

Between October and December 2017, we administered a second follow-up survey ("R2") to 2,151 study households, or 94.0 percent of the baseline sample. In the R2 survey, we did not survey spillover sample households and did not administer language and math tests. Instead, we collected test score data for 649 adolescents who would have been eligible to take the Kenya Certificate of Primary Education (KCPE) examination over the period of the study.

Following Casey, Glennerster, and Miguel (2012), we registered three pre-analysis plans; these are available at http://www.socialscienceregistry.org/trials/350 and in Appendix C. Pre-Analysis Plan A specifies the analyses of the demand and cost data, and Pre-Analysis Plans B and C specify the analyses of electrification impacts using the R1 and R2 survey data, respectively.

C. Data

The analysis combines a variety of survey, experimental, and administrative data, collected and compiled between August 2013 and December 2017. The datasets include: community characteristics data (N=150); baseline household survey data (N=2,504); experimental demand data (N=2,289); administrative community construction cost data (N=77); follow-up household survey data (N=5,696); and children's test score data (N=2,589).

D. Baseline characteristics

[15] In appendix figure B7, we present a timeline of project milestones and grid connection-related news over the study period. Note that by late-2017, a small number of households began to be connected through the LMCP. In 2014, however, neither the sample households nor the research team anticipated such progress. For instance, prior to the intervention, there were concerns that the price would increase; during the intervention, 397 households provided a reason for why they declined a subsidized offer and not one cited the possibility of a lower future price; and the LMCP price reduction was not publicly announced until May 2015, long after subsidy offers had expired. These patterns alleviate concerns that households were anticipating a general price reduction over the course of the experiment.

Table 1 summarizes differences between unconnected and connected households at baseline. Connected households are characterized by higher living standards across almost all proxies for income.[16] They have higher quality walls (made of brick, cement, or stone, rather than mud), have higher monthly basic energy expenditures, and own more land and assets including livestock, household goods (e.g., furniture), and electrical appliances. Most unconnected households (92 percent) rely on kerosene as their primary lighting source, while only 6 and 3 percent of unconnected households own solar lanterns and solar home systems, respectively.

In appendix table B2, we report baseline descriptive statistics and perform randomization checks. On average, 63 percent of respondents are female, just 14 percent have attended secondary school, 66 percent are married, and, in terms of occupation, 77 percent are primarily farmers. These are overwhelmingly poor households, as evidenced by the fact that only 15 percent have high-quality walls. Households have 5.3 members on average. Households spend \$5.55 per month on (non-charcoal) energy sources, primarily kerosene.[17]

We test for balance across treatment arms by regressing baseline household and community characteristics on indicators for the three subsidy levels, and conduct *F*-tests that all treatment coefficients equal zero. For the 23 household-level and two community-level variables analyzed, *F*-statistics are significant at 5 percent for only two variables, namely, a binary variable indicating whether the respondent could correctly identify the presidents of Tanzania, Uganda, and the United States (a measure of political awareness) and monthly (non-charcoal) energy spending, indicating that the randomization created largely comparable groups.

V. Results

A. Estimating the demand for electricity connections

In figure 2, we plot the experimental results on the demand for grid connections. Take-up of a free grid connection offer is nearly universal, but demand falls sharply with price, and is close

---

[16] These patterns are consistent with the stated reasons for why households remain unconnected. In appendix figure B6, we show that, at baseline, 95.5 percent of households cited the high connection price as the primary barrier to connectivity. The second and third most cited reasons—which were the high cost of wiring (10.2 percent) and the high monthly cost (3.6 percent), respectively—are also related to costs. Note that no households said they were unconnected because they were waiting for a lower connection price, or a government-subsidized rural electrification program.

[17] In June 2014, the standard electricity tariff for small households was roughly 2.8 cents per kWh. Taking into consideration fixed charges and other adjustments, \$5.55 translates into roughly 32 kWh of electricity consumption, which is enough for basic lighting, television, and fan appliances each day of the month.

to zero among the low subsidy treatment group, as well as in the control (no subsidy) group. Panel A presents the experimental results and compares them to the government's "prior" on demand, namely, the Ministry of Energy and Petroleum's internal predictions for take-up in rural areas. The government demand curve—which we learned of in early-2015 via a government report—was developed independently of our project and served as justification for the planned LMCP price of \$171 (15,000 KES). A key finding is that, even at generous subsidy levels, actual take-up is significantly lower than predicted by the government (or by our team, see appendix figure B8).[18] In panels B and C, we show that households with high-quality walls and greater earnings in the last month, respectively, had higher take-up rates in the medium and low subsidy arms, suggesting that demand increases at higher incomes.

If we extrapolate the [1.3, 7.1] segment of the demand curve through the intercept, the area under the demand curve is just \$12,421.[19] Based on average community density of 84.7 households, this implies an average valuation of just \$147 per household.

We estimate the following regression equation:

$$y_{ic} = \alpha + \beta_1 T_c^L + \beta_2 T_c^M + \beta_3 T_c^H + X'_c\gamma + X'_{ic}\lambda + \epsilon_{ic} \tag{1}$$

where y_{ic} is an indicator variable reflecting the take-up decision for household i in transformer community c. The binary variables T_c^L, T_c^M, and T_c^H indicate whether community c was randomly assigned into the low, medium, or high subsidy arm, respectively, and the coefficients β_1, β_2, and β_3 capture the subsidy impacts on take-up.[20] Following Bruhn and McKenzie (2009), we include a vector of community-level characteristics, X_c, containing variables used for stratification during randomization (see Section IV.B). We also include a vector of baseline household-level characteristics, X_{ic}, containing pre-specified covariates that may predict take-up (e.g., household size, chickens owned, respondent age, high-quality walls, and whether the respondent attended secondary school, is not a farmer, uses a bank account, engages in business or self-employment, and is a senior citizen). Standard errors are clustered by community, the unit of randomization.

[18] The government report projected take-up in rural areas nationally, rather than in our study region alone, and this is one possible source of the discrepancy. Moreover, the government report does not clearly specify the timeframe over which households would be asked to raise funds for a connection, somewhat complicating the comparison.
[19] In Section V.C, we discuss alternative assumptions regarding demand in the unobserved [0, 1.3] domain.
[20] We focus on this non-parametric specification after rejecting the null hypothesis that the treatment coefficients are linear in the subsidy amount (F-statistic = 23.03), a choice we specified in our pre-analysis plan.

Table 2 summarizes the results of estimating equation 1, where column 1 reports estimates from a model that includes only the treatment indicators, and column 2 includes the household and community controls. All three subsidy levels lead to significant increases in take-up: the 100 percent subsidy increases the likelihood of take-up by roughly 95 percentage points, and the effects of the partial 57 and 29 percent subsidies are much smaller, at 23 and 6 percentage points, respectively. Columns 3 to 8 include interactions between the treatment indicators and household and community characteristics, which are listed in the column headings. Take-up in treatment communities is differentially higher in the low and medium subsidy arms for households with wealthier and more educated respondents; for instance, the coefficient on the interaction between secondary schooling and the medium subsidy indicator is 19.5 percent.[21]

Based on the findings in Bernard and Torero (2015), one might expect take-up to be higher in areas where grid connections are more prevalent if, as they argue, exposure to households with electricity leads individuals to better understand its benefits and value it more. Yet when we include an interaction with the baseline community electrification rate in column 6, or an interaction with the proportion of neighboring households within 200 meters connected to electricity at baseline (column 7), we find no meaningful interaction effects.[22]

B. Estimating the economies of scale in electricity grid extension

Across all projects in the sample, the average total cost per connection ("ATC") is $1,226. While this seems high, it is in line with several alternative estimates, including: (1) Kenya Power's public estimate of $1,435 per rural connection; (2) the Ministry of Energy and Petroleum's estimate of $1,602; and (3) a consultant's estimated range of $1,322 to $1,601 in urban and rural areas, respectively (Korn 2014).[23]

---

[21] In appendix table B3, we compare the characteristics of households choosing to take up electricity across treatment arms. Households that paid more for an electricity connection (i.e., the low subsidy arm) are wealthier on average than those who paid nothing (high subsidy), i.e., they are better educated, more likely to have bank accounts, live in larger households with high-quality walls, spend more on energy, and have more assets. In appendix tables B4A to B4E, we report all related demand regressions specified in our pre-analysis plan, for completeness.

[22] Of course, this does not rule out the possibility of a differential effect at higher levels of electrification, since baseline household electrification rates are generally low in our sample of communities (the interquartile range is 1.8 to 7.8 percent). Also, since community-level characteristics, such as income, are likely positively correlated across households, the lack of statistically significant coefficients may reflect the offsetting joint impacts of negative take-up spillovers and positively correlated take-up decisions; future research could usefully explore these issues.

[23] Elsewhere, rural grid connection costs have been observed to be similar, ranging from $1,100 per connection in Vietnam to $2,300 per connection in Tanzania (Castellano et al. 2015). Note that in our setting, we cannot rule out that connecting a random group of households, rather than a contiguous set of households, may also have increased average costs estimates at low coverage levels.

An immediate consequence of the downward-sloping demand curve estimated above is that the randomized price offers generate exogenous variation in the number of households in a community that are connected as part of the same local construction project. This novel design feature allows us to experimentally assess the economies of scale in grid extension.

In our preferred approach to estimating ATC ($\mathit{\Gamma}_c$) as a function of the number of connections (M_c), we impose the following functional form which features a community-wide fixed cost and linear marginal costs:

$$\mathit{\Gamma}_c = \frac{b_0}{M_c} + b_1 + b_2 M_c \tag{2}$$

Imposing linear marginal costs is both economically intuitive (e.g., as community coverage increases, the marginal cost of connecting an additional household decreases) and closely matches the observed data. Regardless of the exact functional form, average costs decline in the number of households connected, as in the textbook natural monopoly case.[24]

The nonlinear estimation of equation 2 yields coefficient estimates (and standard errors) of b_0= 2,453.4 (s.e. 252.3) for the fixed cost, b_1= 999.4 (s.e. 138.8), and b_2= -3.2 (s.e. 3.6).[25] We take the derivative of the total cost function (which is obtained by multiplying equation 2 by M_c) to estimate the linear marginal cost function:

$$MC_c = b_1 + 2b_2 M_c = 999.4 - 6.5M_c \tag{3}$$

For each community, we use the coefficient estimates to predict the ATC and marginal cost of connecting various levels of community coverage (Q)—defined as the proportion of initially unconnected households in the community that become connected, and which takes on values from 0 to 100. In figure 3, panel A, we compare the experimental demand curve with the ATC and

[24] Note that our preferred nonlinear function differs from the quadratic function specified in our pre-analysis plan. The quadratic function does not provide a good fit to the data: it predicts considerably lower costs at intermediate coverage levels while greatly overstating them at universal coverage. In retrospect, it was an oversight on our part to fail to consider the standard community-level fixed cost. See Section III in Appendix A and appendix figures B9A and B9B for a more detailed discussion on estimating costs and comparisons of different ATC functional forms, respectively.

[25] In Figure 3, we estimate and plot ATC curves by combining two sets of cost data. First, for each community in which the project delivered an electricity connection (n=62), we received budgeted costs for the number of poles and service lines, length of LV lines, and design, labor and transportation costs. We refer to these as "sample" data. Second, REA provided us with budgeted costs for higher levels of coverage (i.e., at 60, 80, and 100 percent of the community connected) for a subset of the high subsidy arm communities (n=15). We refer to these as "designed" data. REA followed the same costing methodology for both (e.g., the same personnel visited the field sites to design the LV network and estimate the costs), ensuring comparability between sample and designed communities. Combining the two sets of communities (N=77) in the main analysis here enables us to trace out ATC across all coverage levels.

marginal cost curves, plotted against Q.[26] Focusing on the ATC curve, we find evidence of strong initial economies of scale. However, the incremental cost savings appear to decline at higher levels of community coverage, and the estimates imply an average cost of approximately \$739 per connection at universal coverage ($Q = 100$).

In communities with larger populations, the higher density of households may potentially translate into a larger impact of scale on ATC. In appendix figures B10A and B10B, we compare ATC curves across various subsamples of data. For instance, appendix figure B10A, panel A, we compare ATC curves for communities with higher and lower populations and find the curves to lie nearly on top of each other. Although it appears there are no significant effects of population on ATC in the range of densities observed in our sample, it seems plausible that ATC could be higher in other parts of Kenya with far lower residential density. In panel B, we compare ATC curves for communities with higher and lower land gradients, and find that while the curves are similar, average cost at universal coverage is somewhat higher for high-gradient communities (at \$839 per connection) compared to low-gradient communities (at \$657 per connection).[27]

C. Experimental approach to estimating social surplus

In figure 3, panel B, we estimate total cost and consumer surplus at full coverage. Note that we first focus on the revealed preference demand estimates, and return to discuss issues of credit constraints and informational asymmetries below in Section VI.

The main observation is that the estimated demand curve for an electricity connection does not intersect the estimated marginal cost curve. To illustrate, at 100 percent coverage, we estimate the total cost of connecting a community to be \$62,618 based on the mean community density of 84.7 households. In contrast, as noted in Section V.A, consumer surplus at this coverage level is far less, at only \$12,421, or less than one quarter the costs. The estimated consumer surplus appears to be substantially smaller than total connection costs at all quantity levels, suggesting that rural household electrification may reduce social surplus. This result is robust to considering the uncertainty in the demand and cost estimates (see appendix figure B9C).

---

[26] Appendix table B5 reports actual and predicted ATC values at various coverage levels.

[27] This result is consistent with Dinkelman (2011), which relies on a positive relationship between land gradient and ATC in South Africa to estimate the impacts of rural electrification on employment. See Section III in Appendix A for a further discussion on the relationship between land gradient and costs.

Specifically, our calculations suggest that a mass electrification program would result in a social surplus loss of $50,197 per community.[28] To justify such a program, discounted future social surplus gains of $593 would be required for each household in the community, above and beyond any economic or other benefits already considered by households in their own private take-up decisions. These social surplus gains could take several possible forms, including spillovers in consumption or broader economic production, an issue we explore below. Credit constraints or imperfect household information about the long-run benefits of electrification may both also contribute to lower demand, while negative pollution externalities could raise social costs.

In an alternative scenario, illustrated in appendix figure B12, we estimate the demand for and costs of a program structured like the LMCP, which planned to offer a connection price of $171. In this case, only 23.7 percent of households would take-up based on the experimental estimates, and thus unless the government were willing to provide additional subsidies or financing, the resulting electrification level would be low. At 23.7 percent coverage, there is an analogous social surplus loss of $18,809 per community, or $935 per connected household.

D. Impacts of rural electrification

Recent literature focuses on estimating the impacts of increasing access to electricity for rural households and communities. However, there is substantial variation in the types of outcomes examined, as well as the magnitudes of impacts estimated.[29] Furthermore, non-experimental studies typically face challenges in identifying credible exogenous sources of variation in electrification status. In contrast, we exploit experimental variation in grid electrification to test the hypothesis that households connected to the electricity grid enjoy improved living standards in the medium-run, roughly 16- and 32-months post-connection.

---

[28] To calculate consumer surplus, we estimate the area under the unobserved [0, 1.3] domain by projecting the slope of the demand curve in the range [1.3, 7.1] through the intercept. The 1.3 percent figure is the proportion of the control group that chose to connect to the grid during the study period, which, for comparability to other points on the demand curve, we assume would happen over the same eight-week period as our offer. If anything, this assumption yields higher consumer surplus than alternative, perhaps more reasonable, assumptions on timing. Appendix figure B11 considers the sensitivity of our results on social surplus loss to alternative demand curve assumptions. In panel C of that figure, the most conservative case, demand is a step function and intersects the vertical axis at $3,000. The social surplus loss is still $39,422 per community in this case.

[29] For example, some studies find that access to electricity increases measures of rural living standards such as income and consumption (Khandker, Barnes, and Samad 2012; Khandker et al. 2014; van de Walle et al. 2015; Chakravorty, Emerick, and Ravago 2016), while others find no evidence of impacts on labor markets outcomes, assets, or housing characteristics (Burlig and Preonas 2016); see Lee, Miguel, and Wolfram (2019) for a discussion.

We limit our discussion of impacts to a set of pre-specified outcomes that are meant to capture several important dimensions of energy access and overall living standards in the study setting.[30] In table 3, we report treatment effects on these outcomes, pooling together R1 and R2 data.[31] Due to relatively low take-up rates in the low and medium subsidy groups, we first limit the sample to include only a comparison between the high subsidy group and the control group and estimate intention-to-treat (ITT) specifications. In column 2, we report the results of estimating the following regression for each outcome:

$$y_{icr} = \beta_0 + \beta_3 T_{Hc} + X_c' \Lambda + Z_{1icr}' \Gamma + s_r + \epsilon_{ic} \tag{4}$$

where y_{icr} represents the primary outcome of interest for household i in community c in round r, T_{Hc} is a binary variable indicating whether community c was randomly assigned into the high-value subsidy treatment, and s_r captures the survey round fixed effect. As in equation 1, we include a vector of community-level characteristics, X_c, as well as a vector of pre-specified, household-level characteristics, Z_{1icr}, and standard errors are clustered at the community level.

We then estimate treatment-on-treated (TOT) results using data from all three subsidy treatment groups. In column 3, we report the results of estimating the following equation:

$$y_{icr} = \beta_0 + \beta_1 E_{icr} + X_c' \Lambda_2 + Z_{1icr}' \Gamma_2 + s_r + \epsilon_{ic} \tag{5}$$

where E_{icr}, is a binary variable reflecting household i's electrification status in round r. We instrument for E_{icr} with the three indicator variables indicating whether community c was randomly assigned to the low, medium, or high subsidy group.

Column 4 reports the false discovery rate (FDR)-adjusted q-values corresponding to the coefficient estimates in column 3, which limit the expected proportion of rejections within a hypothesis that are Type I errors (i.e., false positives).[32]

Energy consumption increases in newly connected households, but overall consumption levels are low. The treatment effect on monthly electricity spending (outcome *A2*, Table 3) is \$1.80 to \$2.17, a miniscule amount corresponding to roughly 2 to 7 kWh of consumption per month. Although kerosene spending (*B7*) decreases by \$0.90 to \$1.00, the effect on total energy spending (*B8*) is much smaller. While there are positive effects to the ownership of certain appliances, such

[30] See Section III in Appendix A and Pre-Analysis Plans B and C for details on the construction of each variable.
[31] Individual survey round results are provided in appendix tables B6A and B6B.
[32] As per our pre-analysis plans, we follow the FDR approach in Casey et al. (2012) and Anderson (2008).

as televisions (*B5*) and irons (*B6*), treated households only modestly expand the number of appliance types owned (*B2*), suggesting that newly connected households use power in limited ways.[33] The vast majority of households in the control group already own mobile phones (85.2%), most own radios (57.6%) and some even own televisions (21.3%).

By the follow-up surveys, there was a small increase in electrification at control households (to 12.2%), partly through the government LMCP in the study region, and a moderate increase in home solar system ownership (to 14.1%). It is noteworthy that, despite the fact that there were major efforts to promote home solar systems in Kenya in this period, and these products are typically available on credit, relatively few control households elected to purchase a system; this may suggest that it is not just a lack of credit that reduces demand for electricity services.

Perhaps surprisingly, but consistent with the results in Section V.B, we do not find evidence of widespread economic or non-economic impacts. There are no detectable effects on asset ownership (*C4*), consumption levels (*C5*), health outcomes (*D1*), or student test scores (*D3*, *D4*). There are moderate and statistically significant impacts on total hours worked (*C3*) and life satisfaction (*D2*), although only the latter is significant at the 5% level when adjusting for multiple testing. The positive life satisfaction effect could reflect a dimension of well-being that we fail to capture in our other primary outcomes, although it could also reflect social desirability bias among respondents. Another possibility is that life satisfaction impacts are transitory since the social status benefits of a grid connection would diminish as more community members are connected.

The overall effects are summarized in table 3, panel E, which combines the primary economic outcomes (*C* outcomes) into a mean effect Economic Index, and primary non-economic outcomes (*D* outcomes) into a Non-Economic Index.[34] The average economic effect is small at 0.02 (in standard deviation units), and reasonably precisely estimated (s.e. 0.06), and the average effect on the non-economic variables is also small at 0.01 (in s.d. units, with s.e. 0.04).[35]

E. Alternative approach to estimating consumer surplus

Alternatively, we can estimate consumer surplus from grid connections using an application of Dubin and McFadden's (1984) discrete-continuous model, similar to Barreca et al.

[33] There are no meaningful impacts to the ownership of other appliance types beyond those presented in table 3.

[34] Although these indices were not pre-specified, they are useful in summarizing the overall results and providing additional statistical power.

[35] As shown in appendix table B6C, we also do not find evidence of any economically meaningful or statistically significant spillover impacts to local households, although these null effects are not precisely estimated.

(2016) and Davis and Killian (2011). This approach allows us to simulate consumer surplus for different cases regarding both baseline consumption levels and long-run consumption growth, under certain assumptions on the functional form of consumer demand.

Households are assumed to make a joint decision to acquire a grid connection and consume electricity, and consumer surplus from the connection is then measured as the discounted sum of surplus from consuming electricity over the life of the connection. We assume zero consumer surplus from electricity without a grid connection.[36] Consumer surplus measures depend on the level of monthly electricity consumption, the demand elasticity for electricity (i.e., the slope of the demand curve), the functional form of the demand curve, the long-run cost of supplying electricity, and the intertemporal discount rate.

This study's experimental variation in grid connection allows us to measure the shift in the demand curve for electricity directly based on connected households' consumption levels. Lacking demand elasticity estimates in Kenya, we use U.S. estimates as a lower bound (e.g., Ito 2014), and report consumer surplus under a range of plausible assumptions. We assume linear demand (following Barreca et al. 2016 and Davis and Killian 2011) with elasticities evaluated at average consumption, a price equal to the constant long-run cost of electricity of \$0.12 per kWh, and an annualized 15 percent discount rate.

Table 4 reports calculated consumer surplus across a range of demand elasticity and consumption cases. In the study sample, mean monthly electricity consumption for newly connected households is 10.8 kWh in R2, an extremely small amount, as noted above. At 10 kWh per month (column 1), consumer surplus ranges from \$98 to \$293 (depending on demand assumptions), and thus falls well below the average connection cost of \$1,226.[37] This result holds even if we assume that energy consumption grows at a rapid 10 percent per year (see column 2); in this case consumer surplus ranges from \$219 to \$658.

Rural connections appear to begin to yield positive social surplus at much higher levels of electricity consumption. Column 3 reports estimates at 70 kWh per month, roughly the mean R2 consumption level reported by households already connected at baseline. Here, consumer surplus

[36] Note that this will, if anything, lead us to overestimate the consumer surplus from acquiring a grid connection since a subset of sample households receive electricity from solar home systems or car batteries.
[37] Note that consumer surplus at the lowest demand elasticity is similar to the average valuation obtained in the experiment, even though we arrive at these figures using two distinct methodologies.

exceeds \$400 (the private cost of a grid connection) and, at low elasticity, rises above the average connection cost in the experiment.[38] Column 4 reports estimates at 190 kWh per month, the mean consumption level in Nairobi.[39] At this level, consumer surplus ranges from \$1,857 to \$5,572.

VI. INTERPRETATION

These results suggesting that rural electrification may reduce social surplus are perhaps surprising. Previous analyses have found substantial benefits from electrification (Dinkelman 2011, Lipscomb, Mobarak, and Barham 2013), though they have not directly compared benefits to costs. In the Philippines, Chakravorty, Emerick, and Ravago (2016) find that the physical cost of grid expansion is recovered after just a single year of realized expenditure gains. A World Bank report argues that household willingness to pay for electricity—which is calculated indirectly based on kerosene lighting expenditures—is likely to be well above the average supply cost in South Asia (World Bank 2008). Most of these studies, however, use non-experimental variation or indirect measures of costs and benefits, and it is possible they do not fully account for unobserved variables correlated with both electrification propensity and improved economic outcomes. In table 1, for example, we document a strong baseline correlation between household connectivity and living standards, and this pattern is consistent with the possibility of meaningful omitted variable bias in some non-experimental studies.

In this section, we consider factors that could boost demand or drive down costs in our setting, affecting the interpretation and external validity of our results. Specifically, we present evidence on the role of credit constraints, bureaucratic red tape, and low grid reliability in reducing demand, and the role of leakage in increasing costs, as well as possibly unaccounted for spillovers.

A. Short-run price reduction and credit constraints

Low demand may be driven in part by household credit constraints, which are well documented in low-income countries (De Mel, McKenzie, and Woodruff 2009; Karlan et al. 2014). In our context, concerns about the role of credit constraints may be exacerbated by the fact that we study a short-run subsidy offer for an electricity connection, redeemable over eight weeks, rather

[38] Note that a full accounting of social surplus for the fraction of households that were initially connected to the grid should include the costs of the transformer and medium-voltage network extensions. Including these would greatly increase the overall costs of rural electrification.
[39] In appendix table B7, we present various benchmarks for monthly electricity consumption throughout Kenya.

than a permanent change in the connection price across villages (which would provide households with more time to raise the necessary funds); long-term differential prices across villages were not feasible in the study setting. This would reduce estimated demand and consumer surplus. On the other hand, short-run subsidies could have the opposite effect: absent credit constraints, temporarily low prices for durables could accelerate purchases from later periods, leading to higher measured willingness to pay (Hendel and Nevo 2006; Mian and Sufi 2012).

In figure 3, panel C, we compare the experimental results to two sets of stated willingness to pay (WTP) results obtained in the baseline survey to shed some light on the possible role of credit constraints. Stated WTP may better capture household valuation in the presence of credit constraints, although they may also overstate actual demand due to wishful thinking or social desirability bias (Hausman 2012).

Respondents were first asked whether they would accept a randomly assigned, hypothetical price ranging from \$0 to \$853 for a grid connection.[40] Households were then asked whether they would accept the hypothetical offer if required to complete the payment in six weeks, a period chosen to be similar to the eight-week payment period in the experiment. We plot results in figure 3, panel C, where the first curve (long-dashed line, black squares) plots the results of the initial question, and the second curve (long-dashed line, grey squares) the follow-up question.

Stated demand is generally high.[41] And, the demand curve falls dramatically when households are faced with a hypothetical time constraint, suggesting they are unable to pay (or borrow) the required funds on relatively short notice, an indication that credit constraints may be binding. At a price of \$171, for example, stated demand is initially 57.6 percent but it drops to 27.2 percent with the time constraint.

Although the experimental demand curve is substantially lower than the stated demand without time limits, it closely tracks the constrained stated demand: at \$171, actual take-up in the experiment is 23.7 percent. The similarity between the constrained stated demand and experimental results suggest that augmenting survey questions to incorporate realistic timeframes

[40] Each of \$114, \$171, \$227, \$284, and \$398 had a 16.7 percent chance of being drawn. Each of \$0 and \$853 had an 8.3 percent chance of being drawn. Nine households are excluded due to errors in administering the question.
[41] For more details on the stated demand for electricity connections, see appendix table B8A, where we estimate the impact of the randomized offers on hypothetical and actual take-up, and appendix table B8B, which includes interactions between indicators for the hypothetical offers and key household covariates.

and other contextual factors could help to elicit responses that more closely resemble revealed preference behavior and are less prone to hypothetical bias (Murphy et al. 2005; Hausman 2012).

We also regressed a binary variable indicating whether a household first accepted the hypothetical offer without the time constraint, but then declined the offer with the time constraint on a set of household covariates. Households with low-quality walls and respondents with no bank accounts are the most likely to switch their stated demand decision when faced with a pressing time constraint, consistent with the likely importance of credit constraints for these groups (see appendix table B8C).[42]

In Section V.C above, we combined the estimated experimental demand and cost curves to show that rural electrification may reduce social surplus. The stated preference results indicate that this outcome is likely to hold even if credit constraints were eased. For example, if we combine the cost curve with the stated demand for grid connections without time constraints, then households in the unobserved [0, 16.7] domain of the stated demand curve (i.e., those willing to pay at least $853) must be willing to pay $2,920 on average for consumer surplus to be larger than total construction costs. While this cannot be ruled out, it appears unlikely in a rural setting where annual per capita income is below $1,000 for most households.[43]

Another way to address credit constraints is to offer financing plans for grid connections. In a second set of baseline stated WTP questions, each household was randomly assigned a hypothetical credit offer consisting of an upfront payment (ranging from $39.80 to $127.93), a monthly payment (from $11.84 to $17.22), and a contract length (either 24 or 36 months); we present the results in figure 3, panel C.[44] Households were first asked whether they would accept the offer (short-dashed line, black circles) and then whether they would accept the offer if required to complete the upfront payment in six weeks (grey circles). We then plot take-up against the net present value of the credit offers based on an annualized 15 percent discount rate.

When households are offered financing, stated demand is not only high but also appears likely to be exaggerated, particularly when there are no time constraints to complete the upfront payments. For example, 52.7 percent of households stated that they would accept the $915.48 net

[42] Relatedly, see appendix figures B13A and B13B for a comparison of hypothetical demand curves for households with and without bank accounts and high-quality walls.
[43] The area under the stated demand curve (without time constraints) is roughly $447 per household, under the assumption that the demand curve can be extended linearly in the [0, 16.7] range, intersecting the y-axis at $2,158.
[44] Results for a range of discount rates and net present values are presented in appendix table B9.

present value offer, a package that consists of an upfront payment of $127.93 and monthly payments of $26.94 for 36 months. Eight weeks after accepting such an offer, a borrower will have paid $181, with an additional $915.92 due in the future. Yet stated demand for this option is twice as high as what we observe for the actual $171 8-week time-limited, all-in price offered to medium subsidy arm households in the experiment. Moreover, the fact that stated take-up is very similar across hypothetical contract offers with quite divergent net present values casts some doubt on the reliability of these stated preference responses. Nonetheless, the area under the stated demand curve in the case with financing and without time constraints is roughly $744 per household (under the same assumptions as above), which again falls short of average costs in our setting.

Figure 3, panel C, combines the four stated demand curves with the experimental demand and ATC curves. Visually, the only demand curves that appear to yield consumer surpluses that are potentially larger than total construction costs are the stated demand curves for grid connections with credit offers, which as we point out above, could be overstated.

Low demand may indicate that even with subsidies, grid connections are simply too expensive for many of the households in our poor rural setting. After the experiment, we asked households that were connected in the low and medium subsidy arms to name any sacrifices they had made to complete their payments: 29 percent of households stated that they had forgone purchases of basic household consumption goods, and 19 percent stated that they had not paid school fees. It seems likely that many households declined the subsidized offer due to binding budget constraints – in other words, poverty – rather than credit constraints alone.

With that said, the ITT results in table 3, column 2 suggest that medium-run impacts of electrification on economic (and other) outcomes are close to zero, even when credit constraints and budget constraints are eliminated by the high subsidy offer, which pushed the connection price to zero. This result implies that consumer surplus from grid connections is likely to be relatively low, unless credit constraints and budget constraints also play a role in limiting appliance purchases and monthly electricity consumption.

B. Other factors contributing to low demand

Low demand may also be partly attributable to the lengthy and bureaucratic process of obtaining an electricity connection. In the experiment, households waited a staggering 188 days on average after submitting their paperwork before they began receiving electricity. The delays were mainly caused by time lags in project design and contracting, as well as in the installation of meters.[45] The World Bank similarly estimates that in practice it takes roughly 110 days to connect new business customers in Kenya (World Bank 2016).

Another major concern is the reliability of power. Electricity shortages and other forms of low grid reliability are well documented in less developed countries (Steinbuks and Foster 2010; Allcott, Collard-Wexler, and O'Connell 2016). In rural Kenya, households experience both short-term blackouts, which last for a few minutes up to several hours, and long-term blackouts, which can last for months and typically stem from technical problems with local transformers. The value a household places on a grid connection could be much lower when service is this unreliable.

During the 14-month period from September 2014 to December 2015 when households were being connected to the grid, we documented the frequency, duration, and primary reason for the long-term blackouts impacting sample communities. In total, 29 out of 150 transformers (19 percent) experienced at least one long-term blackout. On average, these blackouts lasted four months, with the longest lasting an entire year. During these periods, households and businesses did not receive any grid electricity. The most common reasons included transformer burnouts, technical failures, theft, and replaced equipment.[46] As a point of comparison, only 0.2 percent of transformers in California fail over a five-year period, with the average blackout lasting a mere five hours.[47] That said, we find no strong statistical evidence that recent blackouts affect demand: in table 2, column 8, we include interactions between the treatment variables and an indicator for whether any household in the community reported a recent blackout (over the past three days) at baseline, and find no statistically significant effects.

C. Excess costs from leakage

In appendix table B11, we report the breakdown of budgeted versus invoiced electrification costs per community. The budgeted (ex-ante) costs for each project are based on LV network drawings prepared by REA engineers.[48] The invoiced (ex-post) costs are based on actual final invoices submitted by local contractors, detailing the contractor components of the labor, transport,

[45] Field enumerators report that the electricity connection work may have sometimes been delayed due to expectations that bribes would be paid. See Section IV in Appendix A for additional details.
[46] In appendix table B10, we provide a list of all the communities that experienced long-term blackouts.
[47] Based on personal communications with Pacific Gas and Electric Company (PG&E) in December 2015.
[48] An example of an LV network drawing is provided in appendix figure B14.

and materials that were required to complete each project. In total, it cost $585,999 to build 101.6 kilometers of LV lines to connect 478 households through the project.[49]

Overall, budgeted and invoiced costs per connection were nearly identical, amounting to $1,201 and $1,226, respectively. In other words, contractors submitted invoices that were only 1.7 percent higher than the budgeted amount on average.[50] These cost figures reflect the reality of grid extension in rural Kenya. However, it is possible that they are higher than what would ideally be the case due to leakage and other inefficiencies that are common in low-income countries (Reinikka and Svenson 2004). In our context, leakage might occur during the contracting work, in the form of over-reporting labor and transport, which may be hard to verify, and sub-standard construction quality (e.g., using fewer materials than required).[51]

To measure leakage, we sent teams of enumerators to each treatment community to count the number of electricity poles that were installed, and then compared the actual number of poles to the poles included in the project designs and contractor invoices. While there is minimal variation between ex-ante and ex-post total costs, most contractors' projects showed large differences in the number of observed versus budgeted poles with nearly all using fewer poles: the number of observed poles was 21.3 percent less than budgeted, a substantial discrepancy.[52]

Labor and transport costs may also reflect leakage. Labor is typically invoiced based on the number of declared poles, and we show these were inflated. Similarly, transport is invoiced based on the declared mileage of vehicles carrying construction materials. In appendix table B12, we analyze three highly detailed contractor invoices (for nine communities) that we obtained. These data contain evidence of over-reported labor costs associated with the electricity poles, at 11.0 percent higher costs than expected, and over-reported transport costs: based on a comparison

[49] See Section IV in Appendix A for an additional discussion.

[50] The similarity between planned and actual costs provides further confidence that the actual costs for the designed communities (at high coverage levels) would be reasonably accurate (see figure 3).

[51] There is evidence of reallocations across sub-categories in appendix table B11, despite similar ex-ante and ex-post totals. Invoiced labor and transport costs, for example, were 12.7 percent higher than the budgeted amounts, while invoiced local network costs were 6.5 percent lower.

[52] In appendix figure B15, we plot the discrepancies between costs and poles by contractor. In addition to being associated with missing public resources, if the planned number of poles reflects accepted engineering standards (i.e., poles are roughly 50 meters apart, etc.), using fewer poles might lead to substandard service quality and even safety risks. For instance, local households may face greater injury risk due to sagging power lines between poles that are spaced too far apart, and the poles may be at greater risk of falling over. It is possible, however, that REA's designs included extra poles, perhaps anticipating that contractors would not use them all.

between the reported mileage and the travel routes between the REA warehouse and project sites (suggested by Google Maps), invoiced travel costs were 32.9 percent higher than expected.

Taken together, these findings indicate that electric grid construction costs may be substantially inflated due to mismanagement and corruption in Kenya, suggesting that improved contractor performance could reduce costs and possibly improve project quality and safety.[53] On the other hand, note that even with a 20 to 30 percent reduction in construction costs, mass rural household electrification may still lead to a reduction in overall social surplus based on the demand and cost estimates in figure 3, as well as the consumer surplus results in table 4.

D. Factors that increase social surplus from rural electrification

The leading interpretation of our empirical findings is that mass rural household electrification does not lead to greater social surplus in Kenya, according to standard criteria. The cost of electrifying households appears to be five times higher than what households are willing and able to pay for these connections, and consumer surplus appears lower than total costs even when attempting to address credit constraints, or utilizing subsequent electricity consumption patterns among connected households. While per household costs fall sharply with coverage, reflecting the economies of scale in the creation of local grid infrastructure, they appear to remain higher than demand, implying that social surplus falls with each additional subsidized connection. These results are also consistent with the evidence of negligible medium-run economic, health and educational impacts 16- and 32-months post-connection. Further evidence on the low demand for electricity comes from a nearby area in Kenya, where just 1 percent of rural households provided with a large cash transfer of $1,000 chose to connect to the electric grid (Egger et al. 2019).

Yet, it is plausible that these conclusions would change in settings with improved credit markets, better organizational performance by the electricity utility, or different levels of economic development. In table 5, we estimate the social surplus per household using both the experimental approach presented in Section V.C and the alternative demand approach in Section V.E., under a range of assumptions about the underlying institutional and economic setting. In particular, we simulate the impact of "improving" the setting in five distinct ways: (a) allowing for household income growth of 3 percent per annum over 30 years (for the experimental approach) and

[53] To the extent costs are high because contractors are over-billing the government, leakage may simply result in a transfer across Kenyan citizens and not a social surplus loss. The social welfare implications would depend on the relative weight the social planner places on contractors, taxpayers, and rural households.

electricity consumption growth of 10 percent per annum over 30 years (for the alternative approach); (b) alleviating credit constraints for grid connections; (c) eliminating transformer breakdowns; (d) eliminating the connection delays; and (e) eliminating all project construction cost leakage.[54] We examine these individually, and then assess the effect on social surplus of combining them all in what we call the "ideal scenario", which can be thought of as perhaps the best-case scenario for a low-income country considering mass rural residential electrification.

The first row of table 5 presents the base results from the above analysis, including the average connection cost (at 100 percent coverage) of \$739, average consumer surplus from the experimental approach of \$147, and from the alternative approach of \$293. As Kenya continues to develop, it is likely that incomes and energy consumption will grow. To predict the effect of income growth on consumer surplus, we focus on the relative differences between households with low- and high-quality walls. Specifically, we first estimate that households with low-quality walls would need to have income growth of 3 percent annually over ten years in order to reach the income of households with high-quality walls.[55] We then calculate the difference in experimental demand curves between these groups (figure 2, panel B) to be equivalent to a 2.2 percent annual growth rate in consumer surplus over ten years. Extrapolating these relationships over a 30-year period, consumer surplus per household reaches \$285, thus increasing the main estimate of consumer surplus by \$139 (improvement a).

We further refine the estimates of consumer surplus in the experimental approach by relaxing credit constraints, using the valuations from the stated WTP question without time constraints described above (improvement b).[56] This more than triples consumer surplus, but is not enough to alter the conclusion that social surplus is likely to be negative. Similarly, while rapid electricity consumption growth in the coming 30 years (at 10 percent per year) leads to a large increase in consumer surplus in the alternative approach, it is not enough to offset the upfront average connection cost.

---

[54] In appendix table B13, we include an additional adjustment that accounts for the consumer surplus associated with households that were already connected at baseline. This adjustment does not greatly alter our conclusions.

[55] As a proxy for income, we use endline food consumption per capita. Note that we did not have a comprehensive baseline measure of household income or consumption. Our baseline monthly earnings measure—calculated as the sum of respondent profits from businesses and self-employment; salary and benefits from employment; and household agricultural sales—is imperfect as it excludes earnings from other household members as well as subsistence farming.

[56] Note that the alternative approach reflects consumer surplus from a grid connection largely absent credit constraints since it presumes that the household already has a connection.

We next turn to simulated improvements in service provision that address transformer breakdowns (improvement c) and grid connection delays (improvement d), both of which somewhat increase consumer surplus, in the first case by increasing the number of days of service, and in the second case by assuming consumers get access to power sooner. As a rough approximation, we assume demand estimates scale linearly. Neither improvement on its own is sufficiently large enough to overturn the negative social surplus conclusion.

Finally, we simulate a reduction in total construction costs of 21.3 percent consistent with the degree of over-invoicing of construction poles documented in the data (improvement e). This leads to a sharp reduction in total costs under the assumption that this leakage is simply "waste"; leakage would be less socially costly if viewed simply as a transfer from taxpayers to contractors (though would still incur some deadweight loss associated with the cost of raising funds).

The bottom row presents the ideal scenario in which all improvements are simultaneously implemented. The use of the preferred experimental estimates incorporating the easing of credit constraints and future income growth results in a social surplus gain of $83. The alternative estimates using electricity consumption (and assuming rapid future consumption growth) are more positive, with a social surplus gain of $166. The bottom line is that there are optimistic assumptions regarding the reduction of corruption and improvements in electricity service quality, together with sustained economic growth, under which mass rural residential electrification appears to increase social surplus.

There may also be additional benefits that are not captured by household WTP that could make this calculation appear more positive. First, as outlined in Section II.B, there may be spillovers from private grid connections, including any benefits that local unconnected households experience. Yet as mentioned in Section V.A above, we find no evidence of an interaction between the treatment indicators and the local baseline electrification rate.[57] Additionally, as noted in Section V.D, we find no compelling evidence of spillover impacts in R1 data for local unconnected households along a range of economic and non-economic outcomes, although these effects are relatively imprecisely estimated.

[57] Note that we cannot rule out the possibility that any negative effect of these spillovers on take-up due to free-riding is offset by a competing positive "keeping up with the neighbors" mechanism (Bernard and Torero 2015), or that greater learning about the private benefits of electricity and/or correlated household characteristics are present.

Second, grid connections are long-lived but their long-term benefits may not be fully reflected in WTP if households have limited information about the future income or broader social benefits of electrification, or due to imperfect within-household altruism, for instance, if children stand to gain the most from indoor lighting in the evening (if it boosts learnings and future earning) but their parents do not fully understand these gains or incorporate them into decision-making. However, as noted above we do not find evidence for child test score gains in connected households in the medium-run.

Further, other factors may push up costs, making rural electrification less attractive. The per household connection cost would be substantially higher under a policy in which only a subset of households were connected to the grid (given the fixed costs of expanding the local low voltage network), rather than the mass connection case we assume in table 5. Most importantly, access to modern energy could generate negative environmental externalities from higher CO_2 emissions and other forms of pollution.

Finally, we have considered neither the costs nor benefits of the initial investment to extend the high-voltage lines and install transformers in each sample community. Each installation required a relatively large investment—the median cost per transformer is $21,820 (Lee et al. 2016)—and the social surplus gains from powering the targeted public facilities, while potentially large, have not been measured. Our analysis treats these costs as sunk and focuses solely on the economics of electrifying "under grid" households, conditional on existing infrastructure. This is the policy-relevant question in our setting, given the expanding Kenya LMCP, but the cost of transformer installations would need to be considered in many other African and Asian settings.

VII. Conclusion

Over the past century, rural electrification has served as a key benchmark for economic development and social progress. The United States began its mass rural electrification program in the late-1930s, though it required two decades to reach 90 percent of households (Kitchens and Fishback 2015), China did so in the 1950's, and South Africa launched its initiative in the 1990s. Today, access to energy has emerged as a major political issue in many low-income countries.

However, the extent to which increases in energy access should be driven by investments in large-scale infrastructure, such as grid connections, or small-scale decentralized solutions, such as solar lanterns and solar home systems, remains contested. Does Africa's energy future even lie with the grid? Although our findings suggest that rural household electrification may reduce social surplus, they do not necessarily imply that distributed solar systems are any more attractive than the grid, or that the patterns we identify are universal across time and space. In fact, the evidence—on the pervasiveness of bureaucratic red tape, low grid reliability, and household credit constraints, all of which would suppress demand, and inflated construction costs from leakage—suggests that the social surplus consequences of rural electrification are closely tied to organizational performance as well as institutions. We show that settings with better performance by the electricity utility—with fewer losses due to leakage and service that is more responsive to customers—may see shifts in both the cost curve and the demand side, and in such settings mass rural electrification may potentially be socially desirable.

Another possibility is that mass electrification is indeed transformative and reshapes social, political, and economic interactions, perhaps in the long-run, but individual rural households do not internalize these benefits, and they are neither reflected in private demand estimates nor observable in the medium-run follow-up data collected 16- and 32-months post-connection. Rural Kenyan households today may on average be too poor to consume meaningful amounts of electricity, but perhaps after another decade (or two) of sustained income growth they will be able to purchase the complementary appliances needed to fully exploit electrification's promise.

Decisions to invest in large-scale energy infrastructure programs are associated with major opportunity costs and long-run consequences for future economic development and climate change, especially in Sub-Saharan Africa, where access to electricity lags the rest of the world. The findings of this study indicate that connecting rural households today may not necessarily be an economically productive and high return activity in the world's poorest countries. The social returns to investments in transportation, education, health, water, sanitation, or other sectors—indeed possibly including the electrification of industrial sites or urban areas—need to be compared to investments in rural electricity grid expansion to determine the appropriate sequencing of major public investments. Given the high stakes around these decisions, and the limited evidence base, there is a need for research in several areas, including on the impacts of increasing the supply of electricity (both in terms of access and reliability) to different types of consumers, such as commercial and industrial consumers; identifying the patterns and drivers of consumption demand, including for energy-efficient appliances; and determining routes to improving electric utility organizational performance.

References

Abdullah, Sabah, P. Wilner Jeanty. 2011. "Willingness to Pay for Renewable Energy: Evidence from a Contingent Valuation Survey in Kenya." *Renewable and Sustainable Energy Reviews* 15(6): 2974-2983

Allcott, Hunt, Allan Collard-Wexler, Stephen D. O'Connell. 2016. "How Do Electricity Shortages Affect Industry? Evidence from India." *American Economic Review* 106(3): 587-624.

Aker, Jenny C. 2010. "Information from Markets Near and Far: Mobile Phones and Agricultural Markets in Niger." *American Economic Journal: Applied Economics* 2: 46-59.

Anderson, Michael L. 2008. "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(484): 1481-1495.

Barreca, Alan, Karen Clay, Olivier Deschenes, Michael Greenstone and Joseph Shapiro. 2016. "Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century." *Journal of Political Economy* 124(1): 105-159.

Baumol, William J. 1977. "On the Proper Cost Tests for Natural Monopoly in a Multiproduct Industry." *American Economic Review* 67(5): 809-822.

Barron, Manuel, Maximo Torero. 2017. "Household Electrification and Indoor Air Pollution." *Journal of Environmental Economics and Management* 86: 81-92.

Bernard, Tanguy, Maximo Torero. 2015. "Social Interaction Effects and Connection to Electricity: Experimental Evidence from Rural Ethiopia." *Economic Development and Cultural Exchange* 63(3): 459-484.

Bjorkegren, Daniel. 2018. "The Adoption of Network Goods: Evidence from the Spread of Mobile Phones in Rwanda." *Review of Economic Studies* 0: 1-28.

Bruhn, Miriam, David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Econ. Journal: Applied Economics* 1(4): 200-232.

Burlig, Fiona and Louis Preonas. 2016. "Out of the Darkness and Into the Light? Development Effects of Electrification in India." https://ei.haas.berkeley.edu/research/papers/WP268.pdf (accessed April 2019).

Carlton, Dennis W., Jeffrey M. Perloff. 2005. *Modern Industrial Organization*. Boston: Pearson/Addison Wesley.

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.

Castellano, Antonio, Adam Kendall, Mikhail Nikomarov, and Tarryn Swemmer. 2015. "Brighter Africa: The Growth Potential of the Sub-Saharan Electricity Sector." https://www.mckinsey.com/-~/media/McKinsey/dotcom/client_service/EPNG/PDFs/Brighter_Africa-The_growth_potential-_of_the_sub-Saharan_electricity_sector.ashx (accessed April 2019).

Chakravorty, Ujjayant, Kyle Emerick, Majah-Leah Ravago. 2016. "Lighting Up the Last Mile:

The Benefits and Costs of Extending Electricity to the Rural Poor." https://www.rff.org/publications/working-papers/lighting-up-the-last-mile-the-benefits-and-costs-of-extending-electricity-to-the-rural-poor/ (accessed April 2019).

Christensen, Laurits R., and William H. Greene. 1976. "Economies of Scale in U.S. Electric Power Generation." *Journal of Political Economy* 84(4): 655-676.

Davis, Lucas and Lutz Killian. 2011. "The Allocative Cost of Price Ceilings in the U.S. Residential Market for Natural Gas." *Journal of Political Economy* 119 (2): 212-241.

de Mel, Suresh, David McKenzie, Christopher Woodruff. 2009. "Are Women More Credit Constrained? Experimental Evidence on Gender and Microenterprise Returns." *American Economic Journal: Applied Economics* 1(3): 1-32.

Devoto, Florencia, et al. 2012. "Happiness on Tap: Piped Water Adoption in Urban Morocco." *American Economic Journal: Economic Policy* 4(4): 68-99.

Dinkelman, Taryn. 2011. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." *American Economic Review* 101(7): 3078–3108.

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review* 108(4-5): 899-934.

Faber, Benjamin. 2014. "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System." *Review of Economic Studies* 81(3): 1046-1070.

Korn, Andreas. 2014. "Consultancy Services for Development of Electricity Connection Policy and Draft Regulations."

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker. 2019. "General Equilibrium Effects of Cash Transfers in Kenya." Unpublished manuscript, University of California, Berkeley.

Hausman, Jerry. 2012. "Contingent Valuation: From Dubious to Hopeless." *Journal of Economic Perspectives* 26(4): 43-56.

Hendel, Igal, and Aviv Nevo. 2006. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica* 74(6): 1637-1673.

Kitchens, Carl, and Price Fishback. 2015. "Flip the Switch: The Impact of the Rural Electrification Administration 1935-1940." *Journal of Economic History* 75(4): 1161-1195.

IEA (International Energy Agency). 2014. *Africa Energy Outlook*.

Ito, Koichiro. 2014. "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing." *American Economic Review* 104(2): 537-563.

Jensen, Robert. 2007. "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector." *Quarterly Journal of Economics* 122(3): 879-924.

Joskow, Paul. 2000. "Deregulation and Regulatory Reform in the U.S. Electric Power Sector," in *Deregulation of Network Industries: What's Next?* Sam Peltzman and Clifford Winston, eds. Washington, D.C.: AEI-Brookings Joint Center for Regulatory Studies, pp. 113-54.

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, Christopher Udry. 2014. "Agricultural Decisions After Relaxing Credit and Risk Constraints." *Quarterly Journal of Economics* 129(2): 597-652.

Khandker, Shahidur R., Douglas F. Barnes, and Hussain A. Samad. 2012. "The Welfare Impacts of Rural Electrification in Bangladesh." *Energy Journal* 33(1): 187-206.

Khandker, Shahidur R., Hussain A. Samad, Rubaba Ali, and Douglas F. Barnes. 2014. "Who Benefits Most from Rural Electrification? Evidence in India." *Energy Journal* 35(2): 75-96.

Lipscomb, Molly, Mobarak, Ahmed Mushfiq, Tania Barham. 2013. "Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil." *American Economic Journal: Applied Economics* 5(2): 200-231.

Lee, Kenneth, Eric Brewer, Carson Christiano, Francis Meyo, Edward Miguel, Matthew Podolsky, Javier Rosa, Catherine Wolfram. 2016. "Electrification for "Under Grid" Households in Rural Kenya." *Development Engineering* 1: 26-35.

Lee, Kenneth, Edward Miguel, Catherine Wolfram. 2016. "Appliance Ownership and Aspirations among Electric Grid and Home Solar Households in Rural Kenya." *American Economic Review: Papers & Proceedings* 106(5): 89-94.

Lee, Kenneth, Edward Miguel, Catherine Wolfram. 2019. "Does Electrification Supercharge Economic Development?" Unpublished manuscript, University of California, Berkeley.

Mankiw, N. Gregory. 2011. *Principles of Economics, 5th Edition*. Cengage Learning.

Mian, Atif, and Amir Sufi. 2012. "The Effects of Fiscal Stimulus: Evidence from the 2009 Cash for Clunkers Program." *Quarterly Journal of Economics* 127(3): 1107-1142.

Murphy, James J., et al. 2005. "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation." *Environmental and Resource Economics* 30(3): 313-325.

Samuelson, Paul A., and William D. Nordhaus. 1998. *Economics*. Boston: Irwin/McGraw-Hill.

Parshall, Lily, et al. 2009. "National Electricity Planning in Settings with Low Pre-Existing Grid Coverage: Development of a Spatial Model and Case Study of Kenya." *Energy Policy* 37(6): 2395-2410.

Patil, Sumeet R., et al. 2014. "The Effect of India's Total Sanitation Campaign on Defecation Behaviors and Child Health in Rural Madhya Pradesh: A Cluster Randomized Controlled Trial" *PLoS Medicine* 11(8): e1001709.

Reinikka, Ritva, Jakob Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda." *Quarterly Journal of Economics* 119(2): 679-705.

Steinbuks, J., and V. Foster. 2010. "When Do Firms Generate? Evidence on In-House Electricity Supply in Africa." *Energy Economics* 32(3): 505-14.

van de Walle, Dominique, Martin Ravallion, Vibhuti Mendiratta, and Gayatri Koolwal. 2015. "Long-Term Gains from Electrification in Rural India." *World Bank Economic Review*: 1-36.

Viscusi, W. Kip, John M. Vernon, and Joseph Emmett Harrington. 2005. *Economics of Regulation and Antitrust*. Cambridge, Mass: MIT Press.

World Bank. 2008. "The Welfare Impact of Rural Electrification: A Reassessment of the Costs and Benefits. An IEG Impact Evaluation." https://siteresources.worldbank.org/-EXTRURELECT/Resources/full_doc.pdf (accessed April 2019).

World Bank. 2016. "Doing Business 2016: Measuring Regulatory Quality and Efficiency." http://www.doingbusiness.org/content/dam/doingBusiness/media/Annual-Reports/English/-DB16-Full-Report.pdf (accessed April 2019).

Yatchew, Adonis. 2000. "Scale Economies in Electricity Distribution: A Semiparametric Analysis." *Journal of Applied Econometrics* 15: 187-210.

Zvoleff, Alex, Ayse Selin Kocaman, Woonghee Tim Huh, and Vijay Modi. 2009. "The Impact of Geography on Energy Infrastructure Costs." *Energy Policy 37* (10): 4066-4078.

Table 1—Differences between unconnected and grid connected households at baseline

	Unconnected	Connected	p-value of diff.
	(1)	(2)	(3)
Panel A: Household head (respondent) characteristics			
Female (%)	62.9	58.6	0.22
Age (years)	52.3	55.8	< 0.01
Senior citizen (%)	27.5	32.6	0.11
Attended secondary schooling (%)	13.3	45.1	< 0.01
Married (%)	66.0	76.7	< 0.01
Not a farmer (%)	22.5	39.5	< 0.01
Employed (%)	36.1	47.0	< 0.01
Basic political awareness (%)	11.4	36.7	< 0.01
Has bank account (%)	18.3	60.9	< 0.01
Monthly earnings (USD)	16.9	50.6	< 0.01
Panel B: Household characteristics			
Number of members	5.2	5.3	0.76
Youth members (age $\leq$ 18)	3.0	2.6	0.01
High-quality walls (%)	16.0	80.0	< 0.01
Land (acres)	1.9	3.7	< 0.01
Distance to transformer (m)	356.5	350.9	0.58
Monthly (non-charcoal) energy (USD)	5.5	15.4	< 0.01
Panel C: Household assets			
Bednets	2.3	3.4	< 0.01
Sofa pieces	6.0	12.5	< 0.01
Chickens	7.0	14.3	< 0.01
Owns radio (%)	34.8	62.3	< 0.01
Owns television (%)	15.2	80.9	< 0.01
Sample size	2,289	215	

Notes: Columns 1 and 2 report sample means for households that were unconnected and connected at the time of the baseline survey. Column 3 reports p-value of the difference between the means. Basic political awareness indicator captures whether the household head was able to correctly identify the presidents of Tanzania, Uganda, and the United States. Monthly earnings (USD) includes the respondent's profits from businesses and self-employment, salary and benefits from employment, and agricultural sales for the entire household. In the 2013 census of all unconnected households, just 5 percent of rural households were connected to the grid. In our sample of respondents, we oversampled the number of connected households.

Table 2—Impact of grid connection subsidy on take-up of electricity connections

			Interacted variable					
			High-quality walls	Monthly earnings (USD)	Attended secondary school	Baseline electrification rate	Baseline neighbors connected	Report of blackout in past 3 days
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T1: Low subsidy—29% discount	5.8***	5.9***	3.6**	4.8***	4.5***	5.6**	4.8**	6.1**
	(1.4)	(1.5)	(1.5)	(1.5)	(1.4)	(2.2)	(1.9)	(2.6)
T2: Medium subsidy—57% discount	22.4***	22.9***	21.3***	20.9***	19.8***	21.4***	21.4***	18.7***
	(4.0)	(4.0)	(4.4)	(4.1)	(3.8)	(6.2)	(3.5)	(5.1)
T3: High subsidy—100% discount	94.2***	95.0***	95.6***	95.6***	95.2***	97.5***	96.1***	95.1***
	(1.2)	(1.3)	(1.2)	(1.3)	(1.3)	(1.7)	(1.3)	(2.4)
Interacted variable			0.3	-0.0	-1.0	0.1	0.1	-0.9
			(1.4)	(0.0)	(1.5)	(0.1)	(0.1)	(1.3)
T1 × interacted variable			12.3**	0.1*	10.2	0.1	0.2	-0.2
			(6.1)	(0.0)	(7.0)	(0.2)	(0.2)	(3.1)
T2 × interacted variable			8.8	0.1*	19.5***	0.3	0.3	7.6
			(7.8)	(0.1)	(4.6)	(1.2)	(0.2)	(7.8)
T3 × interacted variable			-5.5	-0.0	-4.3	-0.5*	-0.2	-0.2
			(3.9)	(0.0)	(4.9)	(0.3)	(0.1)	(2.8)
Household and community controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,289	2,176	2,176	2,164	2,176	2,176	2,176	2,176
R^2	0.68	0.69	0.69	0.70	0.70	0.69	0.69	0.69

Notes: The dependent variable is an indicator variable (multiplied by 100) for household take-up, with a mean of 21.6. Take-up in the control group is 1.3. Robust standard errors clustered at the community level in parentheses. Pre-specified household controls include the age of the household head, indicators for whether the household respondent attended secondary school, is a senior citizen, is not primarily a farmer, is employed, and has a bank account, an indicator for whether the household has high-quality walls, and the number of chickens (a measure of assets) owned by the household. Pre-specified community controls include indicators for the county, market status, whether the transformer was funded and installed early on (between 2008 and 2010), community electrification rate at baseline, and community population. Monthly earnings (USD) includes the respondent's profits from businesses and self-employment, salary and benefits from employment, and agricultural sales for the entire household. Interacted variables in columns 7 and 8 are the proportion of neighbors (i.e., within 200 meters) connected to electricity and an indicator for whether any households in the community reported a recent blackout, respectively. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table 3—Pooled treatment effects on key outcomes

	Control	ITT	TOT	FDR q-val
	(1)	(2)	(3)	(4)
Panel A: Primary energy outcomes				
A1. Grid connected (%)	12.2	82.8***	–	–
	[32.7]	(1.8)		
A2. Monthly electricity spending (USD)	0.33	1.80***	2.17***	–
	[1.36]	(0.13)	(0.14)	
Panel B: Additional energy outcomes				
B1. Electricity as main lighting source (%)	10.6	72.0***	86.8***	0.001
	[30.8]	(2.1)	(2.1)	
B2. Number of appliance types owned	2.0	0.3***	0.4***	0.002
	[1.4]	(0.1)	(0.1)	
B3. Owns mobile phone (%)	85.2	-2.4	-2.2	0.246
	[35.5]	(1.5)	(1.8)	
B4. Owns radio (%)	57.6	4.6**	7.1***	0.010
	[49.4]	(2.3)	(2.6)	
B5. Owns television (%)	21.3	9.3***	11.6***	0.002
	[40.9]	(2.8)	(3.5)	
B6. Owns iron (%)	5.2	2.9**	3.8***	0.010
	[22.2]	(1.2)	(1.4)	
B7. Monthly kerosene spending (USD)	2.64	-0.90***	-1.00***	0.001
	[2.75]	(0.11)	(0.13)	
B8. Monthly total energy spending (USD)	10.83	-0.36	-0.19	0.870
	[21.83]	(0.99)	(1.18)	
B9. Solar home system as main lighting source (%)	14.1	-13.0***	-16.1***	0.001
	[34.8]	(1.2)	(1.3)	
Panel C: Primary economic outcomes				
C1. Household employed or own business (%)	36.0	2.9	2.2	0.619
	[38.4]	(2.2)	(2.5)	
C2. Per capita monthly household earnings (USD)	12	-1	-2	0.688
	[42]	(2)	(2)	
C3. Total hours worked last week	50.3	-2.6**	-3.5**	0.095
	[24.4]	(1.2)	(1.5)	
C4. Total asset value (USD)	1,237	102	117	0.457
	[1,110]	(76)	(93)	
C5. Per capita consumption of major items (USD)	185	-3	-4	0.721
	[186]	(8)	(9)	

(*Table continued on next page*)

(*Table continued from previous page*)

	Control	ITT	TOT	FDR q-val
	(1)	(2)	(3)	(4)
Panel D: Primary non-economic outcomes				
D1. Recent health symptoms index	0	-0.03	-0.03	0.721
	[1]	(0.06)	(0.07)	
D2. Normalized life satisfaction	0	0.16***	0.19***	0.001
	[1]	(0.04)	(0.04)	
D3. Avg. student test Z-score	0	-0.09	-0.13	0.457
	[1]	(0.09)	(0.10)	
D4. Avg. student KCPE test Z-score	0	-0.12	-0.17	0.550
	[1]	(0.13)	(0.17)	
D5. Political and social awareness index	0	-0.03	-0.01	0.861
	[1]	(0.05)	(0.05)	
D6. Perceptions of security index	0	0.08	0.13*	0.303
	[1]	(0.06)	(0.08)	
Panel E: Mean treatment effects on grouped outcomes				
E1. Economic Index (C outcomes)	0	0.02	0	–
	[1]	(0.06)	(0.07)	
E2. Non-Economic Index (D outcomes)	0	0.01	0	–
	[1]	(0.04)	(0.05)	

Notes: Round 1 and 2 follow-up survey data are pooled together. Column 1 reports mean values in the control group, with standard deviations in brackets. Column 2 reports coefficients from separate ITT regressions in which the dependent variable (e.g., A1) is regressed on the high subsidy treatment indicator. The low and medium subsidy groups are excluded from these regressions. Sample sizes range from 1,419 to 2,894 for these regressions, except for the D3 and D4 regressions, which have sample sizes of 941 and 417, respectively. Column 3 reports coefficients from separate TOT (IV) regressions in which household electrification status is instrumented with the three subsidy treatment indicators. Sample sizes range from 2,094 to 4,295 for these regressions, except for the D3 and D4 regressions, which have sample sizes of 1,411 and 644, respectively. All specifications include pre-specified household, student, and community covariates, as well as a survey round fixed effect. Column 4 reports the FDR-adjusted q-values associated with the coefficient estimates in column 3. FDR-adjusted q-values are computed for each outcome within the additional energy outcomes group (panel B), and for each outcome within the primary outcomes group (panels C and D combined). In panel E, we report mean treatment effects on outcomes grouped into an economic and non-economic index. These groupings were not pre-specified. Robust standard errors clustered at the community level in parentheses. The D4 outcome is the average student z-score on the Kenya Certificate of Primary Education (KCPE) test. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table 4—Alternative approach to estimating consumer surplus per household (HH)

	Monthly electricity consumption / *Benchmark*			
Demand elasticity	10 kWh / *Newly connected HH* (1)	10 kWh / + 10% growth (2)	70 kWh / *Baseline connected HH* (3)	190 kWh / *Nairobi HH* (4)
-0.45	98	219	684	1,857
-0.30	147	329	1,026	2,786
-0.15	293	658	2,053	5,572

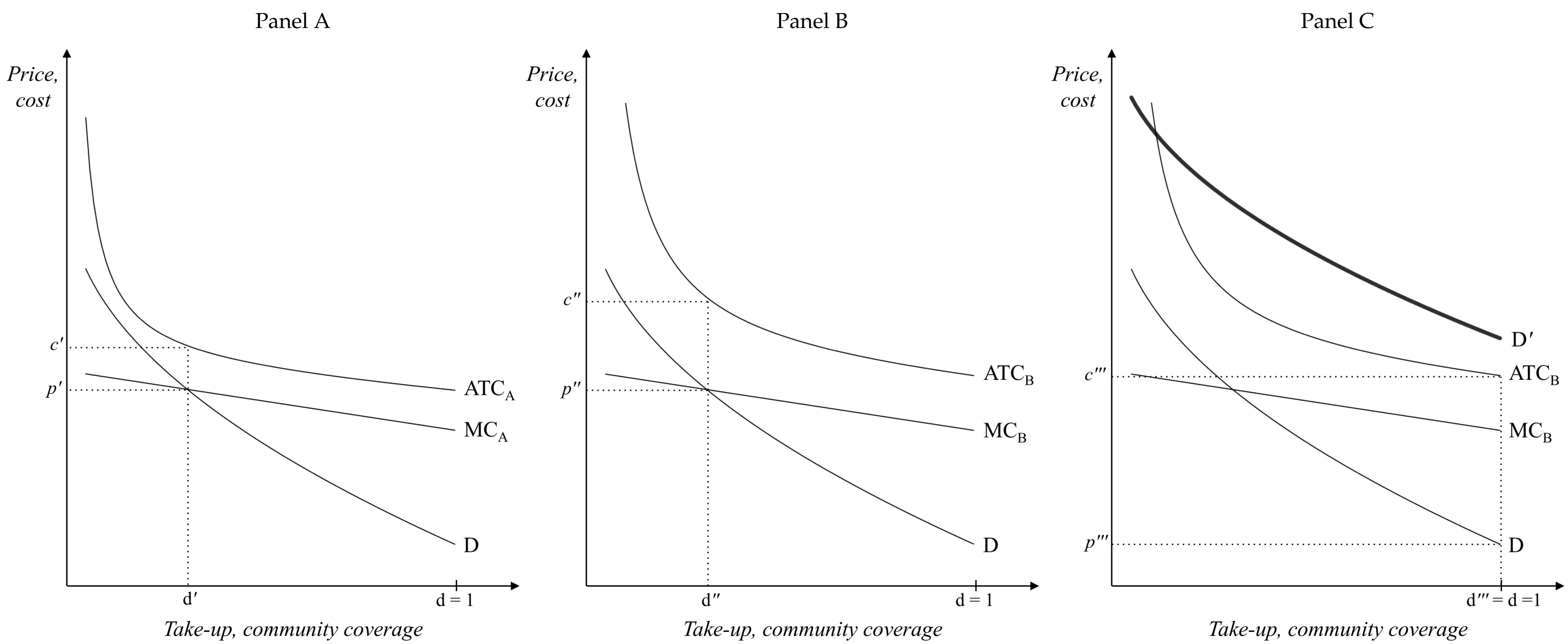
Notes: Consumer surplus is estimated at various monthly electricity consumption levels and consumer demand elasticities. Assumptions include: 15 percent discount rate; 30 year asset life; $0.12 per kWh price; linear demand; zero consumer surplus from electricity without a grid connection; 188 day connection delay. Mean consumption levels are: 10.8 kWh for newly connected HHs in R2; 72.3 kWh for baseline connected HHs in R2; 189.9 kWh for Nairobi HHs in 2014. See appendix table B7 for additional benchmarks.

Table 5—Predicting social surplus per household (SS) under different assumptions

		Experimental approach		Alternative approach		
	C	CS	SS	CS	SS	Key assumption(s)
Main estimates	739	147	-593	293	-446	
a) Income growth (*experimental approach*)	–	+139		–		Growth of 3 percent per annum over 30 years (based on figure 2, panel B).
Electricity consumption growth (*alternative approach*)	–	–		+365		Growth of 10 percent per annum over 30 years (see table 4, column 2, row 3).
b) No credit constraints for grid connections	–	+301		–		Stated WTP without time constraints (see figure 3, panel C)
c) No transformer breakdowns	–	+33		+37		Reduce transformer breakdowns from 5.4 to 0 percent (see appendix table B10).
d) No connection delays	–	+46		+52		Reduce waiting period from 188 to 0 days (see appendix figure A1).
e) No construction cost leakage	-157	–		–		Decrease total construction costs by 21.3 percent (see appendix table B11).
Ideal scenario	582	665	83	747	166	

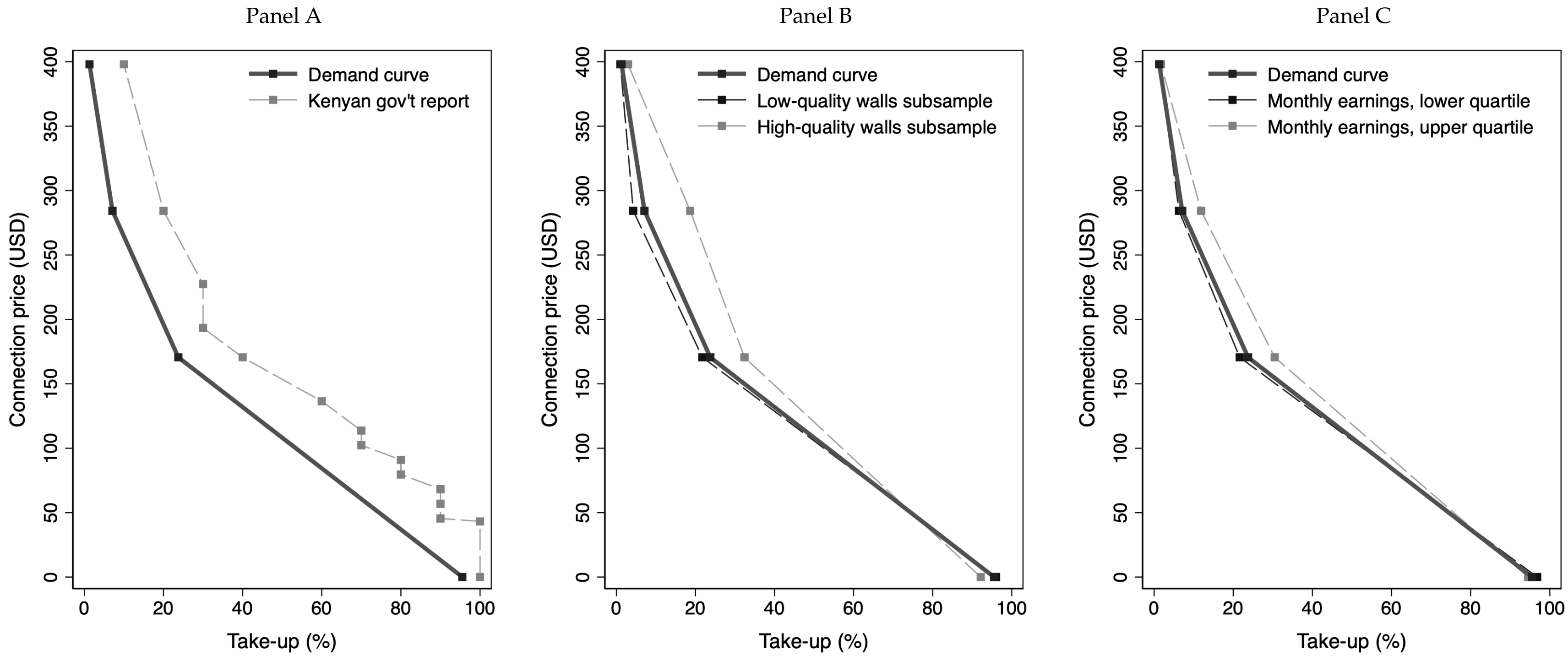
Notes: Main estimates of C, CS, and SS correspond to figure 3, panel B (for the experimental approach), and table 4, column 1, row 3 (for the alternative approach). Appendix table B13 includes an additional row to account for the consumer surplus associated with baseline connected households.

Figure 1—The electric utility as a natural monopoly



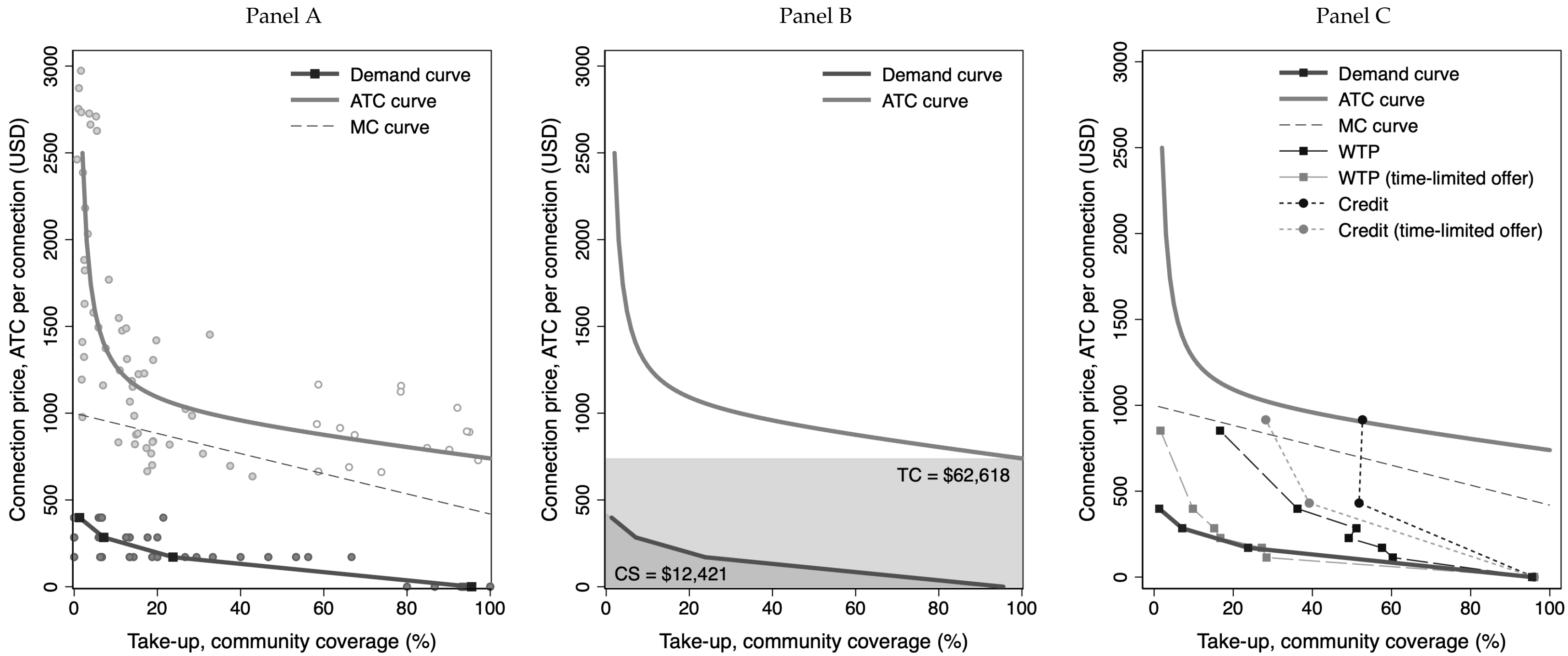
Notes: In panel A, the electric utility is a natural monopoly facing high fixed costs, decreasing marginal costs (MC_A), and decreasing average total costs (ATC_A). MC_A intersects demand at d′. At d′, a government-subsidized mass electrification program would increase social surplus since consumer surplus (i.e., the area under the demand curve) is greater than total cost. Panel B illustrates an alternative scenario with higher fixed costs. In this case, consumer surplus is less than total cost at all quantities. A mass electrification program would not increase social surplus unless there are, for instance, positive externalities from private grid connections. Panel C illustrates a scenario in which social demand (D′) is sufficiently high for the ideal outcome to be full coverage, subsidized by the government.

Figure 2—Experimental evidence on the demand for rural electrification



Notes: Panel A compares the experimental results to the assumptions in an internal government report shared with our team in early-2015. Panel B plots the experimental results separately for households with low- and high-quality walls. Panel C plots the results separately for households in the lower and upper quartiles of monthly earnings, which is defined as the respondent's profits from businesses and self-employment, salary and benefits from employment, and agricultural sales for the entire household.

Figure 3—Experimental evidence on the social surplus implications of rural electrification



Notes: Panel A combines the experimental demand curve with the population-weighted average total cost per connection (ATC) curve corresponding to the predicted cost of connecting various population shares, based on the nonlinear estimation of $ATC = b_0/M + b_1 + b_2M$. Each point represents the community-level, budgeted estimate of ATC at a specific level of coverage. Panel B demonstrates that the estimated total cost of community electrification is $62,618, based on average community density of 84.7 households. The area under the demand curve is estimated to be $12,421. These estimates suggest that a mass electrification program would result in a social surplus loss of $50,197 per community (i.e., $593 per household). Panel C combines the curves in panel A with the contingent valuation (CV) questions included in the baseline survey. The CV questions included: (1) whether the household would accept a hypothetical offer (i.e., at a randomly assigned price) to connect to the grid; (2) whether the household would accept the same offer if required to complete the payment in six weeks. The credit offer consisted of an upfront payment (ranging from $39.80 to $79.60), a monthly payment (ranging from $11.84 to $17.22), and a contract length (either 24 or 36 months). We plot the net present value of the credit offers, assuming a 15 percent discount rate. Additional details on the credit offers are provided in appendix table B9.

Supplementary Appendix for Online Publication

Experimental Evidence on the Economics of Rural Electrification

Kenneth Lee, Edward Miguel, and Catherine Wolfram

April 2019

Appendix A

I. Theoretical Framework

A representative household will decide to connect to the electricity grid if the benefits from future electricity consumption minus the cost of that consumption exceed the cost of the connection. We represent those tradeoffs formally with the following equation, which reflects the household utility as a function of grid connection status. The indicator *G* equals one if the household connects and zero if not:

$$V(G) = \begin{cases} E\left[\sum_{t=1}^{T} \beta^t \left(Max_{d_t}\, u(y(d_t), x(d_t)) - p_t d_t\right)\right] - CP, if\ G = 1 \\ 0, if\ G = 0 \end{cases} \tag{1}$$

Discounted expected future household utility is denoted by *V*. For simplicity, we normalize household utility in the absence of a grid connection to zero ($G = 0$). If a household connects to the grid in period $t = 0$, it must pay the connection price $CP \geq 0$. In each period $t = 1$ to T (i.e., the lifetime of a connection), the household chooses a level of electricity consumption d_t to maximize the difference between the per-period utility benefits of electricity (u), and the cost $p_t\, d_t$. Under standard assumptions, household electricity demand is a decreasing function of the contemporaneous electricity price p_t, which we assume is linear in consumption for simplicity. Note that we are ignoring dynamic considerations in per-period electricity consumption decisions, although this could be incorporated with more notation. We are also assuming that the household will enjoy the connection for its whole lifetime. Equation 1 demonstrates that household expectations regarding future electricity prices and future consumption factor into the upfront grid connection decision problem.

Households benefit from electricity both in terms of economic outcomes (denoted y) and non-economic outcomes (x), both of which are presumed to be weakly increasing in consumption d_t. Households may have poor information regarding the magnitude of these future benefits if they have not experienced an electric connection themselves. $\beta < 1$ is the time discount factor.

The expression in equation 1 is equal to private consumer surplus from an electricity connection. It also equals the social surplus benefits of electricity connections under additional assumptions. Specifically, the sum of consumer surplus across households is equivalent to the net social surplus benefit if: the cost of a unit of electricity once connected is equal to the marginal cost of supply; there is a perfectly elastic supply curve; and there are no spillover effects, externalities or additional costs from either electricity connections or consumption.

It is useful to extend the expression in equation 1 to consider some of these factors, namely, the possibility that there are spillovers, external effects or additional costs. In the real-world application we study in Kenya, the connection price faced by households (*CP* in the above) was heavily subsidized in all cases, and thus a household's decision to connect imposes a further social cost of $C_0 \geq 0$, which captures the subsidy the household receives. Additional electricity consumption may also impose negative externalities on others to the extent that the marginal cost of supply does not incorporate broader social costs of electricity generation, for example pollution from electricity generation. This cost per unit of electricity is denoted $s_t \geq 0$.[1] Greater energy consumption could also generate positive externalities for other households, denoted *b*, to the extent that there are agglomeration economies, economies of scale, direct spillovers (e.g., neighbors visit each other to watch TV), or other forms of production complementarities across households. *b* could also capture within-family benefits, for example, if parents make decisions about grid connections without fully internalizing the future benefits to their children's earning capacities.

Taking these factors into account, the social surplus that results from a household's decision to connect to the electricity grid can be represented as follows:

$$SW(G) = \begin{cases} E\left[\sum_{t=1}^{T} \beta^t \left(u(y(d_t), x(d_t)) + b(d_t) - (p_t + s_t)d_t\right)\right] - (CP + C_0), if\ G = 1 \\ 0, if\ G = 0 \end{cases} \tag{2}$$

The connection decision (*G*) and the per-period electricity consumption levels (d_t) here are determined by the household's private optimization problem from equation 1, and thus may not be socially optimal in the presence of the additional costs and spillover terms.

The terms in this expression are closely linked to the empirical estimates in the current study. The estimated revealed preference of household willingness to pay for an electricity connection (in Section V.A) captures whether households expect that the price of a connection

[1] Note that we are assuming the firm providing electricity faces a zero-profit constraint, for instance, because it is regulated. In other words, we are assuming that $\sum_t \sum_i \left(p_t d_{it} - mc(d_{it})\right) + \sum_i (CP + C_0) - F = 0$, where $mc(\ )$ is the firm's marginal cost function, *F* represent its fixed costs and $\sum_i (\ )$ sums over the firm's customers, *i*. C_0 reflects transfers from the government (or multilateral development banks). This assumption simplifies the social surplus calculations, and the firm is not the focus of our analysis.

(CP) is less than the discounted future stream of utility benefits minus the expected costs of electricity consumption, as represented by the first expression in equation 1. The alternative measures of surplus from grid connections using the application of Dubin and McFadden's (1984) discrete-continuous model (in Section V.E) utilizes per-period household electricity consumption levels combined with assumptions regarding the elasticity of consumer demand to derive the net present value of consumer surplus. This is essentially measuring $u(\,) - pd$ each period and taking the discounted sum over the assumed lifetime of the connection.

In Section V.D, we present estimates of the medium-run impacts of a grid connection along both economic (y) and non-economic (x) dimensions. In addition, we present estimates of local spillovers (b) in the appendix. Note that the spillover estimates we present would not capture any benefits that accrue to households beyond the contemporaneous village-level impacts. Finally, the cost estimates in Section V.B provide estimates of $CP + C_0$.

II. Experimental Design and Data

A. Sample selection

In August 2013, REA representatives in Western Kenya provided us with a master list of 241 unique REA projects, consisting of roughly 370 individual transformers spread across the ten constituencies of Busia and Siaya. Since REA had been the main driver of rural electrification, this master list reflected the universe of rural communities in which there was a possibility of connecting to the grid. Each project featured the electrification of a major public facility (market, secondary school, or health clinic), and involved a different combination of high and low voltage lines and transformers. Projects that were either too recent, or classified as "not commissioned," were not included in the master list. Since the primary objective was to estimate local electrification rates, projects that were funded after February 2013 were excluded to ensure that households in sample communities had had ample opportunity to connect to the grid.

In September 2013, we randomly selected 150 transformers using the following procedure: 1) in each constituency, individual transformers were listed in a random order, 2) the transformer with the highest ranking in each constituency was then selected into the study, and 3) any remaining transformers located less than 1.6 km (or 1 mile) from, or belonging to the same REA project, as one of the selected transformers, were then dropped from the remaining list. We repeated this procedure, cycling through all ten constituencies, until we were left with a sample of

150 transformers for which: 1) the distance between any two transformers was at least 1.6 km, and 2) each transformer represented a unique REA project. In the final sample, there are 85 and 65 transformers in Busia and Siaya counties, respectively, with the number of transformers in each of the ten constituencies ranging from 8 to 23. This variation can be attributed to differences across constituencies in the number of eligible projects. In Budalangi constituency, for example, all of the eight eligible projects were included in the sample. As a result of this community selection procedure, the sample is broadly representative of the types of rural communities targeted by REA in rural Western Kenya.

B. Experimental design and implementation

1. Households were identified at the level of the residential compound, which is a unit known locally as a *boma*. In Western Kenya, it is common for related families to live in different households within the same compound.
2. Most of the baseline surveys were conducted between February and May 2014. However, 3.1 percent of surveys were administered between June and August 2014 due to scheduling conflicts and delays.
3. Since electrification rates were so low, the sample of connected households covers only 102 transformer communities; 17 communities did not have any connected households at the time of census, and we were unable to enroll any connected households in the remaining 31 communities, for instance, if there was a single connected compound in a village and the residents were not present on the day of the baseline survey.
4. For the stratification variable market status, we used a binary variable indicating whether the total number of businesses in the community was strictly greater than the community-level mean across the entire sample.
5. To prevent transfers of the connection offer between households, the offer was only valid for the primary residential structure, identified by the GPS coordinates captured during the baseline survey. All treatment households were given a reminder phone call two weeks prior to the expiry date of the offer. At the end of the eight-week period, enumerators visited each household to collect copies of bank receipts to verify that payments had been made.

C. Data

The analysis combines a variety of survey, experimental, and administrative data, collected and compiled between August 2013 and December 2017. The datasets include:

1. *Community characteristics data* (N=150) covering all 150 transformer communities in our sample, including estimates of community population (i.e., within 600 meters of a central transformer), baseline electrification rates, year of community electrification (i.e., transformer installation), distance to REA warehouse, and average land gradient. (Following Dinkelman (2011), gradient data is from the 90-meter Shuttle Radar Topography Mission (SRTM) Global Digital Elevation Model (www.landcover.org). Gradient is measured in degrees from 0 (flat) to 90 (vertical).)
2. *Baseline household survey data* (N=2,504) consisting of respondent and household characteristics, living standards, energy consumption, and stated demand (contingent valuation) for an electricity connection.
3. *Experimental demand data* (N=2,289) consisting of take-up decisions for the 1,139 treatment households (collected between May and August 2014) and 1,150 control households (collected between January and March 2015) in our sample.
4. *Administrative cost data* (N=77) supplied by REA including both the budgeted and invoiced costs for each project. For each community in which the project delivered an electricity connection (n=62), we received data on the number of poles and service lines, length of LV lines, and design, labor and transportation costs. Using these data, we calculate the average total cost per household for each community. In addition, REA provided us with cost estimates for higher levels of coverage (i.e., at 60, 80, and 100 percent of the community connected) for a subset of the high subsidy arm communities (n=15). (REA followed the same costing methodology, e.g., the same personnel visited the field sites to design the LV network and estimate the costs, applied to the communities in which we delivered an electricity connection, to ensure comparability between budgeted estimates for "sample" and "designed" communities.) Combining the actual sample and designed communities data (N=77) enables us to trace out the cost curve at all coverage levels.

5. *Round 1 follow-up household survey data* (N=3,545) consisting of respondent and household characteristics, living standards, energy consumption, and other variables, roughly 16 months after treatment households were connected.
6. *Round 2 follow-up household survey data* (N=2,151) consisting of respondent and household characteristics, living standards, energy consumption, and other variables, roughly 32 months after treatment households were connected.
7. *Children's test score data* (N=2,951) consisting of: (a) standardized scores on an English and Math test, administered during the Round 1 follow-up survey (n=2,302); and (b) standardized scores on the national Kenya Certificate of Primary Education (KCPE) examination, collected during the Round 2 survey (n=649).

III. ADDITIONAL RESULTS

A. Estimating the economies of scale in electricity grid extension

An immediate consequence of the downward-sloping demand curve estimated in Section V.A is that the randomized price offers generate exogenous variation in the number of households in a community that are connected as part of the same local construction project. This novel design feature allows us to experimentally assess the economies of scale in electricity grid extension.

In appendix tables A1A, A1B, and A1C, we report the results of estimating the impact of the number of connections (M_c) and a quadratic term (M_c^2)—or alternatively, the impact of the community coverage (Q_c) and a quadratic term (Q_c^2)—on the average total cost per connection ("ATC") (Γ_c). Community coverage is defined as the proportion of initially unconnected households in the community that become connected. For example, for the number of connections, we estimate the following regression:

$$\Gamma_c = \pi_0 + \pi_1 M_c + \pi_2 M_c^2 + V_c' \mu + \eta_c \qquad (2)$$

In the pre-analysis plan, we hypothesized that the ATC would fall with more connections (i.e. $\pi_1 < 0$), but at a diminishing rate (i.e. $\pi_2 > 0$). We test this using two samples. The first sample consists of the 62 treatment communities in which we observed non-zero demand. The second sample includes the additional 15 sites that were designed and budgeted for us by REA at even higher coverage levels (up to 100 percent) (see footnote 25). We report the results for the “sample” communities in appendix table A1A, and “sample and designed” communities in appendix table A1B. In certain columns, we report the coefficients for the community-level characteristics specified in the pre-analysis plan, including for instance, the round-trip distance between community c and the regional REA warehouse in Kisumu (a determinant of project transport costs), and the average land gradient for each 600-meter radius transformer community. In appendix table A1A, columns 5 to 8, we report the results of an instrumental variables specification in which the experimental subsidy terms, T_c^M and T_c^H serve as instruments for either the number of connections (M_c and M_c^2) or community coverage (Q_c and Q_c^2).[2]

The coefficients on M and M^2 are both statistically significant and large with the hypothesized signs. In addition, we find no evidence of endogeneity as the OLS and IV estimates are quite similar. Within the domain of the first sample (appendix table A1A), which ranges from 1 to 16 connections per community, increasing project scale by a single household decreases the ATC by roughly \$500, and costs reach a minimum at approximately 11 households. Within the domain of the second sample (appendix table A1B), which includes the designed communities and ranges from 1 to 85 connections, the estimated π_1 drops to roughly \$84 and costs reach a minimum at approximately 55 households.

In appendix table A1B, column 3, we estimate the ATC as a quadratic function of community coverage, Q_c, We carry out this transformation (focusing on Q_c instead of M_c) because estimating the ATC in terms of community coverage will allow for a direct comparison of the demand curve to the cost curves in Section V.C. Using the estimated coefficients, we predict the cost of connecting various population shares for each community, and then plot the population-weighted ATC curve in appendix figure B9A, panel A. The quadratic function does not provide a particularly good fit to the data: it predicts considerably lower costs at intermediate coverage levels while greatly overstating them at universal coverage. Given this pattern, and the conceptual importance of community-level fixed costs, it thus appears preferable to estimate ATC using the nonlinear functional form that accounts for community-level fixed costs.

In appendix table A5C, column 2, we include interactions between scale (i.e., number of connections) and community population. In line with our observations in Section V.B, we find no

[2] In our pre-analysis plan, we specified an IV regression that included three instrumented variables, M_c, M_c^2, and M_c^3. We dropped the third term because we were unable to acquire cost estimates for the control communities, which limited our sample to the treatment communities, and effectively limited our set of instruments to T_c^M and T_c^H.

significant effects of community population on ATC in the range of densities observed in our sample. Yet it seems plausible that per household connection costs could be higher in other parts of rural Kenya with far lower rates of residential density (see appendix table B1).

In appendix table A5C, column 3, we include interactions between scale and land gradient. In contrast to our observations in Section V.B, we find no evidence that higher average land gradient is associated with higher ATC. Although this result is perhaps counterintuitive, it is important to note that there is little variation in average land gradient in our sample, which ranges from 0.79 to 7.76 degrees. Land gradient may be an important predictor of the costs associated with extending high-voltage lines to new areas in KwaZulu-Natal, South Africa, as in the Dinkelman (2011) case. Our data suggest, however, that it may be less important in predicting the costs of grid connections across smaller areas, at least in our setting.

B. Key outcome variable definitions

In Table 3, we report the results of estimating the impacts of grid connections on a set of pre-specified outcomes that are meant to capture several important dimensions of energy access and overall living standards in the study setting. The outcome variables are as follows:

Panel A: Primary energy outcomes

A1. Grid connected (%): Indicator for whether the household is connected to the electricity grid.

A2. Monthly electricity spending (USD): Total spending over the past month on prepaid and postpaid electricity bills.

Panel B: Additional energy outcomes

B1. Electricity as main lighting source (%): Indicator for whether electricity is identified as the household's main lighting source.

B2. Number of appliance types owned.

B3. Owns mobile phone (%): Indicator for whether the household owns a mobile phone.

B4. Owns radio (%): Indicator for whether the household owns a radio.

B5. Owns television (%): Indicator for whether the household owns a television.

B6. Owns iron (%): Indicator for whether the household owns an iron.

B7. Monthly kerosene spending (USD): Total spending over the past month on kerosene.

B8. Monthly total energy spending (USD): Total spending over the past month on all energy sources, including electricity, kerosene, batteries, fuel, firewood, etc.

B9. Solar home system as main lighting source (%): Indicator for whether a solar home system is identified as the household's main lighting source.

Panel C: Primary economic outcomes

C1. Household employed or own business (%): Proportion of household members (18 and over) currently employed or running their own business.

C2. Per capita monthly household earnings (USD): Sum of earnings for all household members that are employed or running their own business, as well as agricultural earnings, divided by the number of household members.

C3. Total hours worked last week: Total hours worked in agriculture, self-employment, employment, and household chores in the past seven days.

C4. Total asset value (USD): Estimated value of savings, livestock, electrical appliances, and other assets, valued at local market prices.

C5. Per capita consumption of major items (USD): Estimated value of annual consumption of 23 common household goods, divided by the number of household members.

Panel D: Primary non-economic outcomes

D1. Recent health symptoms index: Standardized index of health symptoms (e.g., fever, persistent cough, stomach pain, etc.) experienced by the respondent over the past four weeks. Higher scores reflect better health (i.e., less symptoms experienced).

D2. Normalized life satisfaction: Standardized self-reported life satisfaction (i.e., happiness on a scale of 1 to 10).

D3. Avg. student test Z-score: Average standardized performance on an English and Math test administered to children between the ages of 12 to 16 in R1.

D4. Avg. student KCPE test Z-score: Average standardized performance on the KCPE (Kenya Certificate of Primary Education) test collected for students between the ages of 13 and 17 in R2.

D5. Political and social awareness index: Standardized performance by the respondent on a series of questions about current events. Higher scores reflect better knowledge.

D6. Perceptions of security index: Index of crime experienced by the respondent over the last 12 months. Higher scores reflect better security (i.e., less crime experienced).

Panel E: Mean treatment effects on grouped outcomes

E1. Economic Index (C outcomes): Index of primary economic outcomes in panel C.

E2. Non-Economic Index (D outcomes): Index of primary non-economic outcomes in panel D.

IV. INTERPRETATION

A. Factors contributing to lower demand for electricity connections

In our sample, households waited a staggering 188 days, after submitting all their paperwork, before they began receiving electricity. Appendix figure A1 summarizes the time required to complete each major phase associated with obtaining a rural household grid connection in Kenya. The timeline is presented in two panels; panel A reflects the experience of households, and panel B reflects supplier performance. (In appendix table A2, we document the full list of reasons for the delays encountered during each phase.) From the household's perspective, we identified three phases in the connection process: Payment (A1), Wiring (which also includes submitting a metering application to Kenya Power) (A2), and Waiting (A3).

Unexpected delays occurred during the wiring phase, which on average took 24 days, for two main reasons. First, households applying to Kenya Power are required to have (1) a National Identity Card (NIC), (2) a KRA Personal Identification Number (PIN) certificate, and (3) a completed Kenya Power application form. Forty-two percent of household heads requesting a connection did not already have a KRA PIN certificate, which could only be generated on the KRA website. Since most rural households do not regularly access the Internet, project enumerators provided registration assistance for 96.6 percent of the households lacking KRA PINs. At the time of the experiment, KRA PIN registration services were typically offered at local Internet cafes at a cost of \$5.69 (500 KES). Second, households connecting to the grid are required to have certificates that the wiring is safe. The ready-board manufacturer provided wiring certificates that needed to be signed by contractors after installation. We encountered delays when the spelling of the name on the certificate did not precisely match its spelling on the NIC or KRA PIN certificate.

From the supplier's perspective, we identified four phases: Design (B1), Contracting (B2), Construction (B3), and Metering (B4). REA completed the design and contracting work, independent contractors (hired by REA) completed the physical construction, and Kenya Power educated households on issues relating to safety, and installed and activated the prepaid meters. The longest delays occurred during the design phase, which took an average of 57 days, and the metering phase, which took 68 days on average. The design phase was adversely affected by competing priorities at REA. In June 2014, the government announced a program to provide free laptops for all Primary Standard 1 students nationwide. Since roughly half of Kenya's primary schools were unelectrified at the time of the announcement, there was political pressure on REA to prioritize connecting the remaining unelectrified primary schools during the 2014-15 fiscal year. As a result, fewer REA designers were available to focus on other projects, including ours.

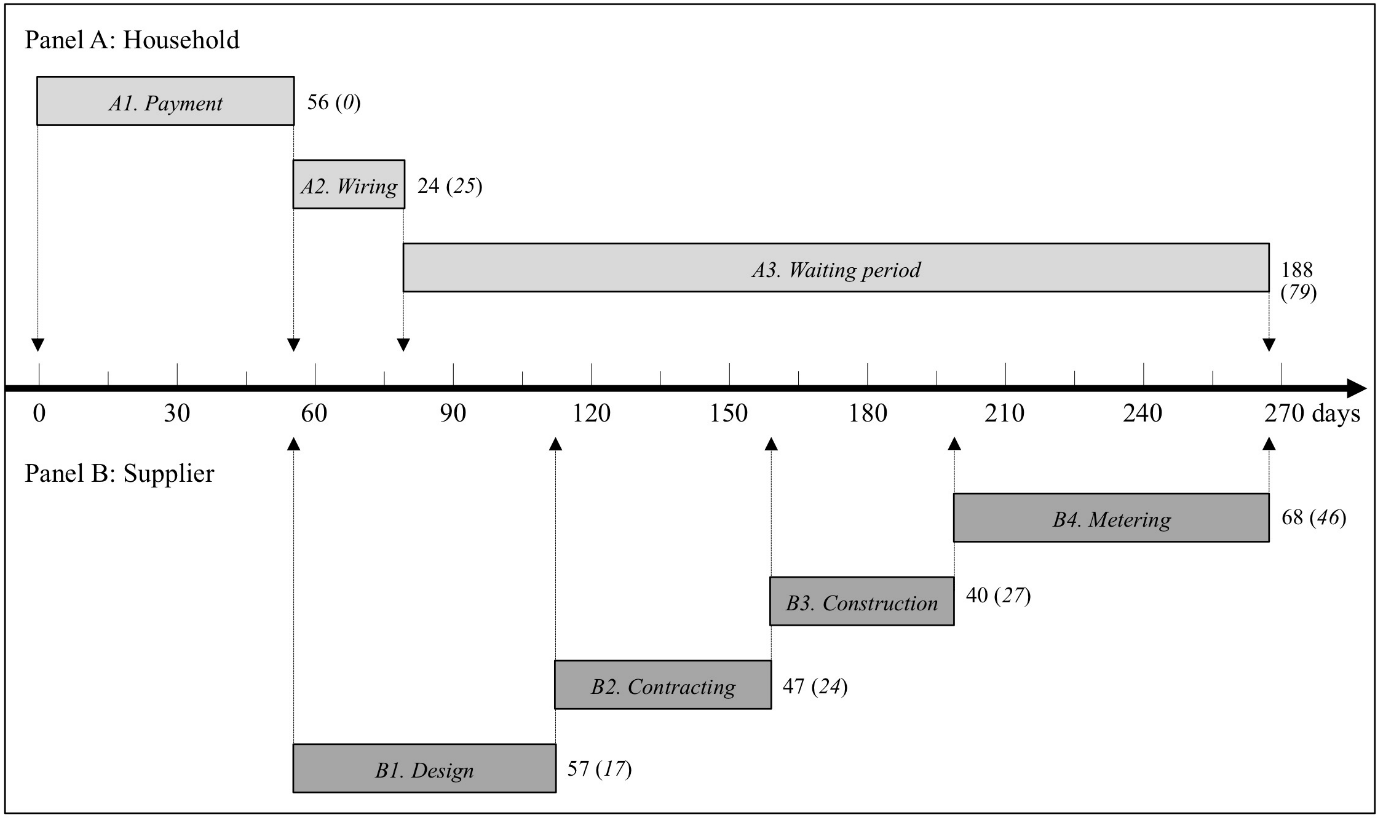
There were severe delays during the metering phase due to unexpected issues at Kenya Power, such as insufficient materials (i.e., reported shortages in prepaid meters), lost meter applications, and competing priorities for Kenya Power staff. Additional problems slowed the process as well. For several months, there was a general shortage of construction materials and metering hardware at REA storehouses. In the more remote communities, heavy rains created impassable roads. Difficulties in obtaining wayleaves (i.e., permission to pass electricity lines through other private properties) required redrawing network designs, additional trips to the storehouse, and further negotiations with contractors. In some cases, households that had initially declined a "ready board" changed their minds; in an unfortunate case lightning struck, damaging a household's electrical equipment; and so on. While these problems increased completion times, their negative effects were partially offset by the weekly and persistent reminders sent to REA and Kenya Power by our project staff, meaning the situation for other rural Kenyans could be even worse.

B. Excess costs from leakage

In addition to being associated with wasted public resources, if the planned number of poles reflects accepted engineering standards (i.e., poles are roughly 50 meters apart, etc.), using fewer poles might lead to substandard service quality and even safety risks. For instance, local households may face greater injury risk due to sagging power lines between poles that are spaced too far apart, and the poles could be at greater risk of falling over. It is possible, however, that

REA's designs included extra poles, perhaps anticipating that contractors would not use them all. We separate costs into three categories: (1) *Local network costs*, which consist of low- and high-voltage cables, wooden poles and the various components required to attach cables to poles, (2) *Labor and transport costs*, which include the cost of network design, installation, and transportation, and (3) *Service lines*, which are the drop-down cables connecting the homes. Note that in appendix table B11, we exclude the costs of metering (incurred by Kenya Power) and ready-boards. Including them would not alter the main conclusions since they are the same for all connected households and a small share of total costs.

Figure A1—Timeline of the rural electrification process



Notes: Panel A summarizes the rural electrification process from the standpoint of the household, divided into three key phases. Panel B summarizes the process from the standpoint of the supplier, divided into four key phases. The numbers to the right of each bar report the average number of days required to complete each phase (standard deviations in parantheses). Households were first given 56 days (8 weeks) to complete their payments. Afterwards, it took on average 212 days (7 months) for households to be metered and electricity to flow to the household. Appendix table A2 lists specific issues that created delays during each phase of the process.

Table A1A—Impact of scale on average total cost per connection (ATC), sample communities

	Sample—OLS				Sample—IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of connections (M)	-472.4***	-510.1***			-551.6**	-492.5**		
	(87.3)	(95.6)			(257.8)	(247.8)		
M^2	20.4***	23.2***			25.0*	22.1		
	(4.6)	(5.2)			(14.8)	(14.2)		
Community coverage (Q)			-177.0***	-171.8***			-409.7**	-335.0*
			(27.3)	(30.8)			(205.5)	(171.2)
Q^2			3.2***	3.0***			11.7	9.3
			(0.7)	(0.8)			(7.5)	(6.3)
Busia=1		583.8		470.8		574.0*		966.9
		(352.0)		(375.1)		(333.4)		(765.1)
Market transformer=1		-342.1*		-190.8		-332.9		-375.9
		(189.8)		(191.1)		(217.0)		(380.5)
Transformer funded early on=1		85.2		114.9		85.9		-136.4
		(188.1)		(214.5)		(174.3)		(381.4)
Community electrification rate		3.4		14.0		3.6		14.9
		(13.6)		(14.6)		(13.3)		(17.4)
Community population		-0.3		-1.4**		-0.4		-0.2
		(0.5)		(0.7)		(0.5)		(1.2)
Round-trip distance to REA (km)		-2.5		-0.5		-2.4		-6.9
		(3.9)		(4.1)		(4.1)		(9.5)
Land gradient		-153.2**		-136.5		-152.3**		-107.6
		(67.8)		(89.5)		(59.6)		(154.1)
Mean of dep. variable (USD)	1,813	1,813	1,813	1,813	1,813	1,813	1,813	1,813
Observations	62	62	62	62	62	62	62	62
R^2	0.63	0.71	0.54	0.62	–	–	–	–

Notes: The dependent variable is the budgeted average total cost per connection (ATC) in USD. Community coverage (Q) is the proportion of unconnected households that are connected (multiplied by 100). Since there was no takeup in 13 communities, there are 62 observations. In columns 5 to 8, polynomials for the number of connections (M and M^2) and community coverage (Q and Q^2) are instrumented with T^M and T^H. Specifications in columns 2, 4, 6, and 8 include (and report coefficients for) the pre-specified community-level covariates. Robust standard errors in parantheses. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table A1B—Impact of scale on average total cost per connection (ATC), sample and designed communities

	Sample & Designed—OLS			
	(1)	(2)	(3)	(4)
Number of connections (M)	-87.8***	-81.1***		
	(15.1)	(16.5)		
M^2	0.8***	0.8***		
	(0.2)	(0.2)		
Community coverage (Q)			-84.3***	-84.6***
			(12.5)	(13.3)
Q^2			0.8***	0.8***
			(0.1)	(0.1)
Busia=1		247.7		487.7
		(388.8)		(361.7)
Market transformer=1		-148.8		-153.3
		(195.4)		(177.8)
Transformer funded early on=1		109.3		240.0
		(218.6)		(193.7)
Community electrification rate		15.9		15.5
		(15.4)		(14.6)
Community population		-0.7		-1.2*
		(0.7)		(0.6)
Round-trip distance to REA (km)		1.6		-1.7
		(3.6)		(3.2)
Land gradient		-173.9***		-186.5***
		(58.1)		(66.6)
Mean of dep. variable (USD)	1,633	1,633	1,633	1,633
Observations	77	77	77	77
R^2	0.43	0.48	0.47	0.55

Notes: The dependent variable is the budgeted average total cost per connection (ATC) in USD. Community coverage (Q) is the proportion of unconnected households that are connected (multiplied by 100). The sample is expanded to include the 15 additional designed communities. The specifications in columns 2 and 4 include (and report coefficients for) the community-level covariates specified in the pre-analysis plan. Robust standard errors in parantheses. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table A1C—Impact of scale on average total cost per connection (ATC), sample and designed communities

	Sample & Designed—OLS			
	(1)	(2)	(3)	(4)
Number of connections (M)	-81.1***	-96.7***	-83.4***	-109.2***
	(16.5)	(18.0)	(17.0)	(18.6)
M^2	0.8***	1.0***	0.8***	1.3***
	(0.2)	(0.2)	(0.2)	(0.2)
Community population		-0.5		
		(1.0)		
Community population × M		0.0		
		(0.1)		
Community population × M^2 / 100		-0.1		
		(0.1)		
Land gradient			-599.3***	
			(164.1)	
Land gradient × M			36.7**	
			(13.9)	
Land gradient × M^2			-0.3*	
			(0.2)	
Households				-4.9
				(11.3)
Households × M				0.1
				(0.5)
Households × M^2 / 100				-0.9*
				(0.5)
Community controls	Yes	Yes	Yes	Yes
Mean of dep. variable (USD)	1,633	1,633	1,633	1,633
Observations	77	77	77	77
R^2	0.48	0.52	0.54	0.55

Notes: The dependent variable is the budgeted average total cost per connection (ATC) in USD. The dataset includes both sample and designed communities. Column 1 displays the same results as column 2 in appendix table A1B. Average land gradient ranges from 0.79 to 7.76 degrees with a mean of 2.15 degrees. Column 4 includes interaction terms for the (demeaned) number of households (i.e., residential compounds) in each community. Note that this variable is not included in the standard list of controls. Robust standard errors in parantheses. All specifications include the pre-specified community-level covariates. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table A2—Reasons for unexpected delays in household electrification

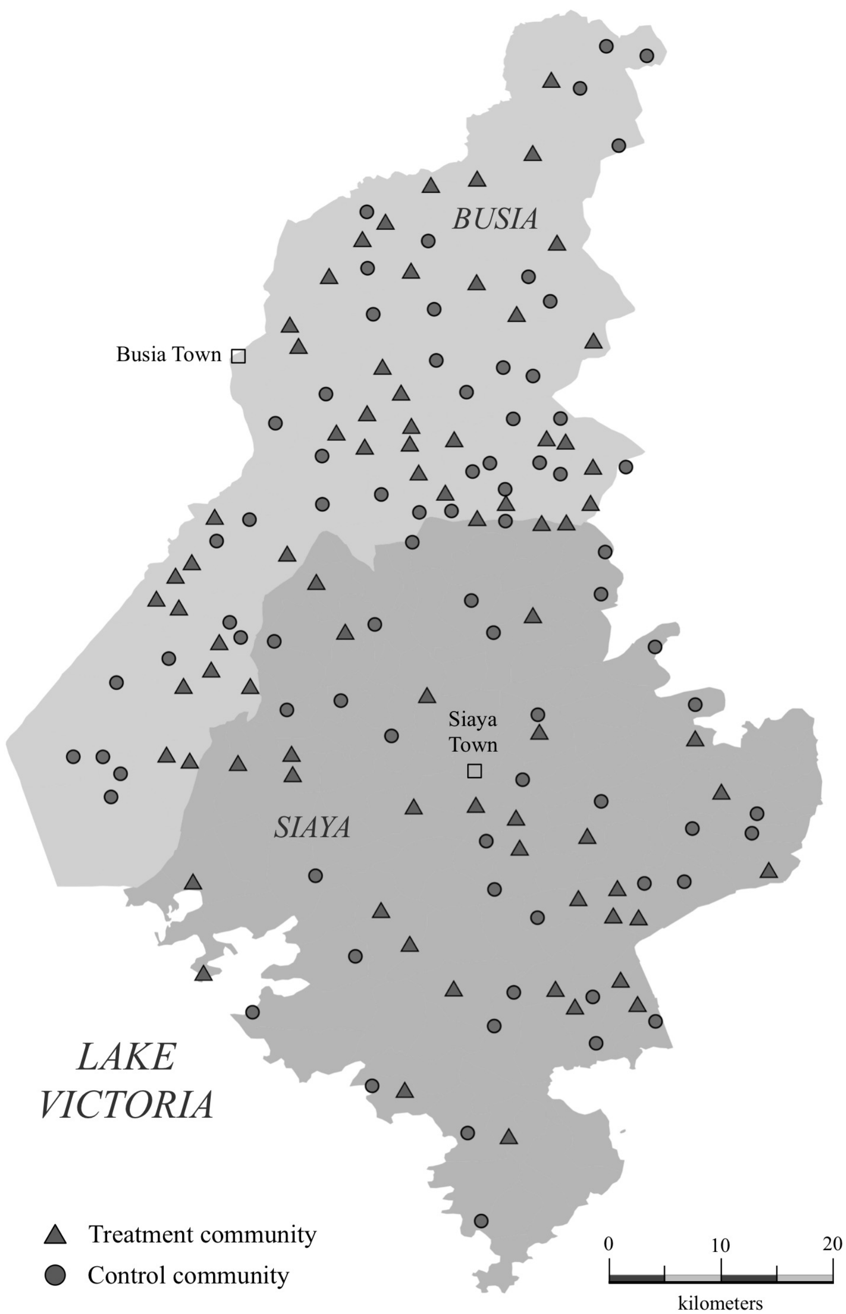
Phase	Description	Reasons for unexpected delays
A2	Wiring	• In order to begin using electricity, households are required to have a valid meter and a certificate of wiring safety. A large proportion of households were not able to register for a meter because they lacked a PIN (*Personal Identification Number*) certificate from the Kenya Revenue Authority. In our sample, 42 percent of households applying for electricity needed assistance in applying for a PIN certificate.
B1	Design	• Competing priorities at REA due to the 2014/15 nationwide initiative to connect primary schools to the national grid. This resulted in a persistent shortage of REA designers and planners. • Low motivation to perform design duties. In addition, since REA designers were required to physically visit each community, there were numerous challenges in scheduling field visits.
B2	Contracting	• Competing priorities (described above) delayed the bureaucratic paperwork required to prepare contracts. • REA staff members had strong preferences to assign certain projects to specific contractors. This resulted in delays because REA wanted to wait until specific contractors were free to take on new projects.
B3	Construction	• Insufficient materials (e.g., poles, cables) requiring site revisits. • Poor weather (i.e., rainy conditions) made roads impassable and digging holes (for electricity poles) impossible. • Issues in securing wayleaves (i.e., right of ways) to pass through neighboring properties. • Low-quality construction work that needed to be fixed. • Missing materials. • Faulty transformers requiring contractors to revisit sites to complete the final step of the process (e.g., connecting the new low-voltage network to the existing line). • Incorrect households were connected to the network, requiring site revisits. • Contractor issues installing "ready-boards" due to lack of experience.
B4	Metering	• Insufficient materials (e.g., prepaid meters, cables) contributed to lengthy delays at Kenya Power. • Lost meter application forms at local Kenya Power offices. • Changes in internal Kenya Power processes requiring applications to be approved in Nairobi as well as local offices in Siaya, Kisumu, and Busia. • Unexpected requests by local Kenya Power representatives for additional documents (e.g., photocopies of payment receipts). • Local Kenya Power representatives unable to perform metering duties due to competing priorities. • Scheduling difficulties due to the necessity for Kenya Power to make multiple trips to remote village sites, which increased the costs (metering costs are not documented in our cost estimates).

Notes: Each phase of the construction process corresponds to the timeline bar illustrated in appendix figure A1.

Appendix B

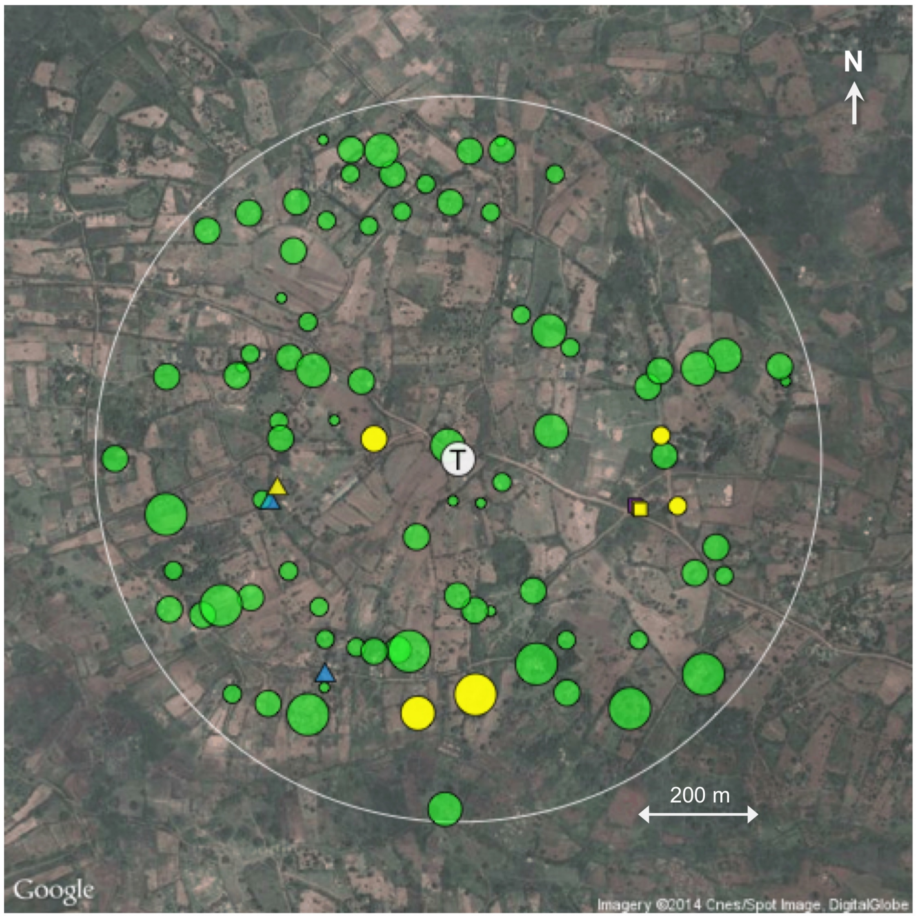
This appendix contains additional figures and tables referenced in the main text.

Figure B1—150 sample communities in Busia and Siaya counties in Kenya



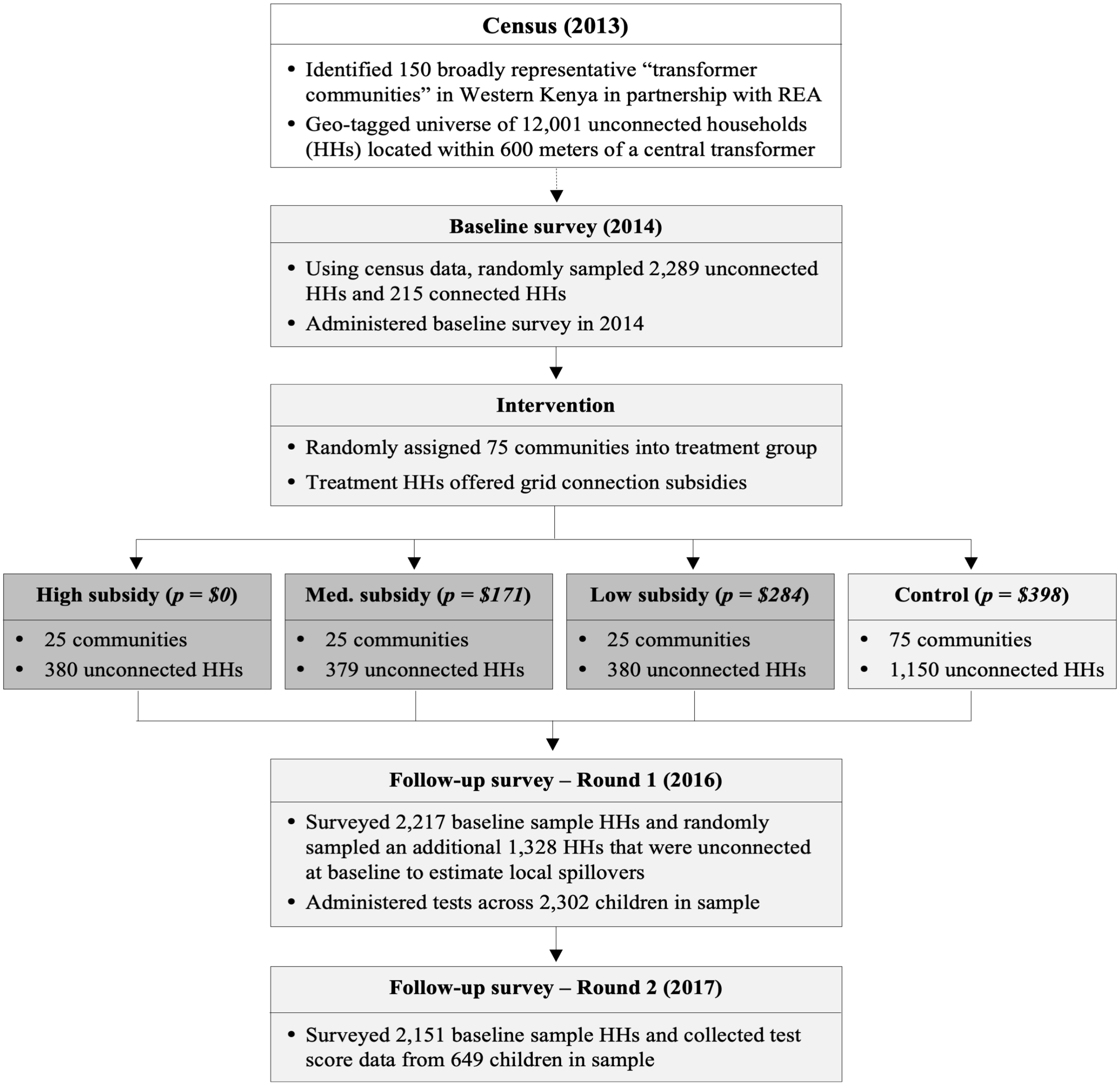
Notes: The final sample of 150 communities includes 85 and 65 transformers in Busia and Siaya counties, respectively.

Figure B2—Example of a "transformer community" of typical density



Notes: The white circle labeled T in the center identifies the location of the REA transformer. The larger white outline demarcates the 600-meter radius boundary. Green circles represent unconnected households; purple squares represent unconnected businesses; and blue triangles represent unconnected public facilities. Yellow circles, squares, and triangles indicate households, businesses, and public facilities with visible electricity connections, respectively. Household markers are scaled by household size, with the largest indicating households with more than ten members, and the smallest indicating single-member households. In each community, roughly 15 households were randomly sampled and enrolled into the study. The average density of a transformer community is 84.7 households per community and the average minimum distance between buildings (i.e., households, businesses, or public facilities) is 52.8 meters. In the illustrated community, there are 85 households.

Figure B3—Experimental design



Notes: The 150 transformer communities in our sample covered 62.2 percent of the universe of REA projects in Busia and Siaya counties in August 2013. See appendix A for details on the community selection procedure. At baseline, roughly 15 unconnected households in each community were randomly sampled and enrolled into the study. Census data on the universe of unconnected households were used as a sampling frame. Baseline surveys were also administered to a random sample of 215 households already connected at baseline. Communities were randomly assigned into three treatment arms and a control group. Treatment offers were valid for eight weeks. In the first follow-up survey, roughly nine additional households in each community were randomly sampled and enrolled into the study in order to measure local spillovers. Census data on the universe of unconnected households were again used as a sampling frame.

Figure B4—Example of REA offer letter for a subsidized household electricity connection

Rural Electrification Authority

ipa INNOVATIONS FOR POVERTY ACTION

NYANZA REGION OFFICE
Mamboleo junction, Opposite Lake Basin Development Authority, Kisumu – Kakamega Highway
P.O. Box 2604 – 40100

Tom Mboya Drive,
Milimani Kisumu.
P.O. Box 313

ELECTRICITY SUPPLY TO YOUR PREMISES (ESP)

UNIQUE CUSTOMER LOCATION DETAILS:

Project name: |______________________| Reference number: |______________________|

Transformer: |______________________| Substation number: |______________________|

Customer name: |______________________________________|

Customer address: |______________________________________|

Household coordinates: |______________________| LATITUDE

|______________________| LONGITUDE

Dear Sir/Madam,

Reference is made to the ongoing *Rural Electric Power Project* research project that is being carried out by Innovations for Poverty Action (IPA) in partnership with REA in this community. As part of this research project, we are pleased to advise you that a single-phase service cable can be installed to your premises at the following price:

SINGLE-PHASE SUPPLY CONNECTION: KES [EFFECTIVE PRICE]

If you would like to accept this offer, kindly arrange to pay this amount to the Rural Electrification Authority account at Kenya Commercial Bank, Milimani Branch, Account Number: 1103201557.

NOTICE: When making your payment, please remember to quote your unique ID number: [NUMBER] in the memo line. Please keep a copy of your bank payment slip for official receipting. This receipt is very important and must be shown to IPA in order to complete this process. No payment should be made to any individual. Your payment should only be made to the REA bank account above. This offer is valid for 8 weeks:

DATE OF OFFER: DD-MM-YY **DATE OF EXPIRY:** DD-MM-YY

For further enquiries, please contact the REA Regional Coordinator, Nyanza Region, [NAME] at [MOBILE] or the IPA Project Associate in Busia Town, [NAME] at [MOBILE].

For: RURAL ELECTRIFICATION AUTHORITY

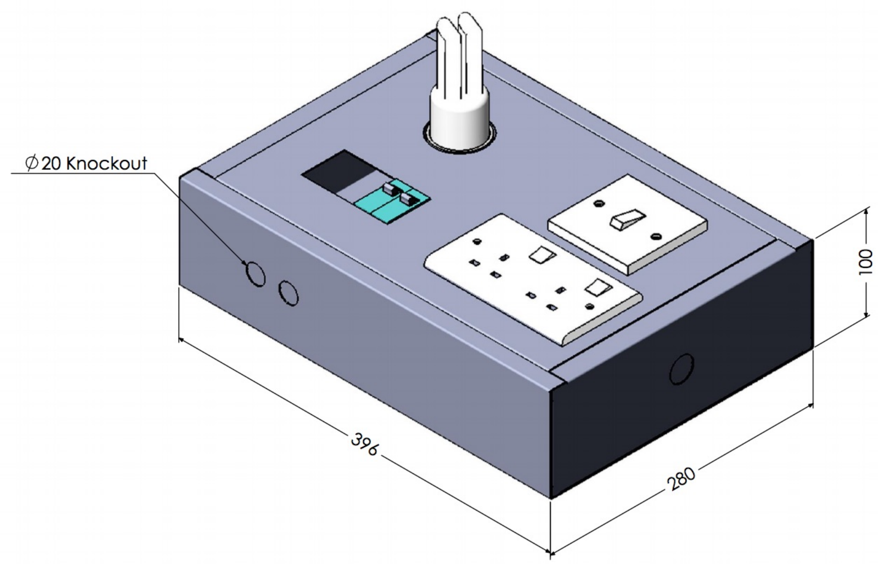
____

[NAME]
REGIONAL CO-ORDINATOR, NYANZA REGION.

Unique ID: [NUMBER]

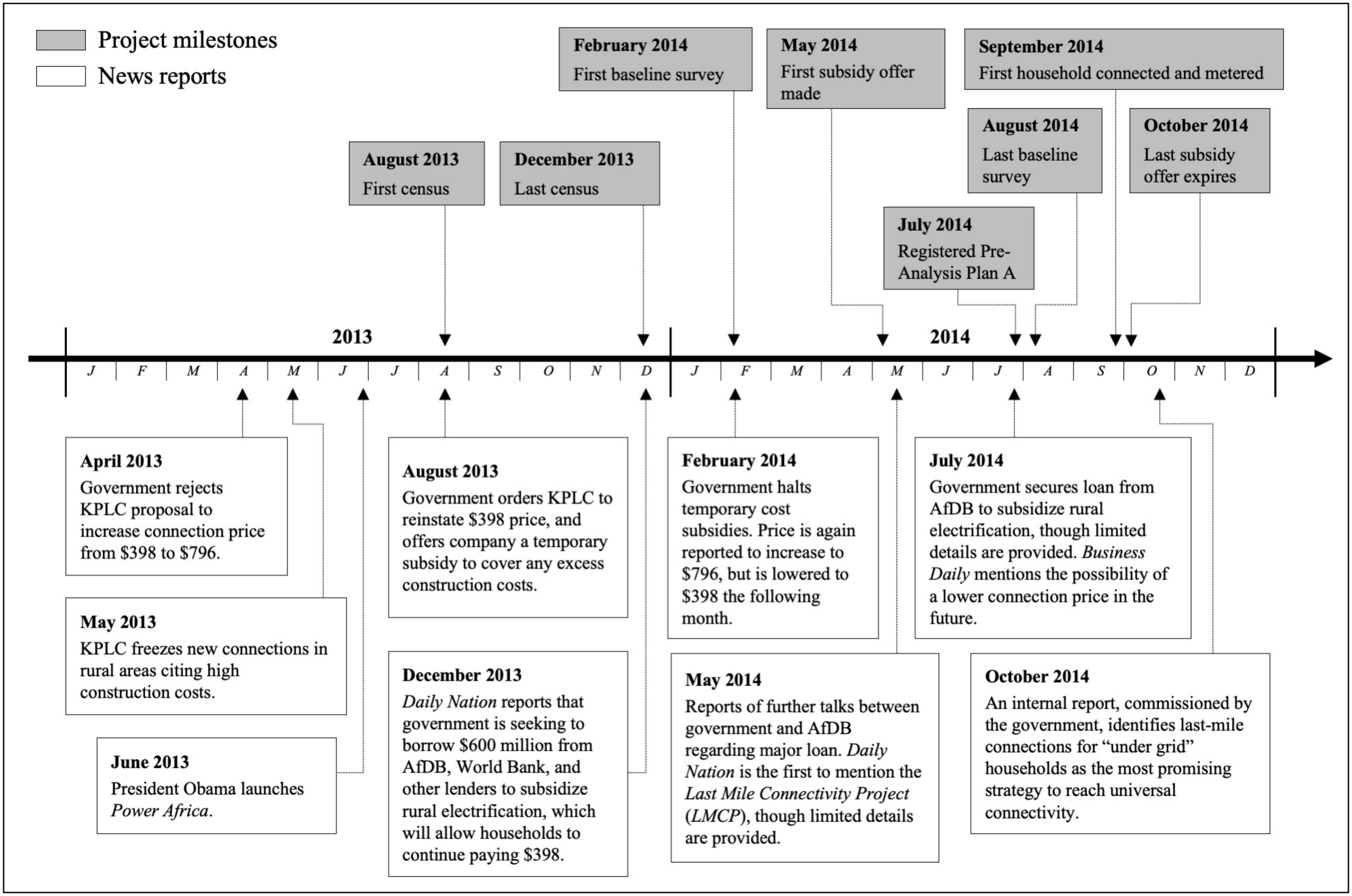
Notes: Each offer letter was signed and guaranteed by REA management. Project field staff members visited each treatment community and explained the details of the offer to a representative from each household in a community meeting. The meeting was held to give community members a chance to ask questions.

Figure B5—*Umeme Rahisi* "ready-board" designed by Power Technics



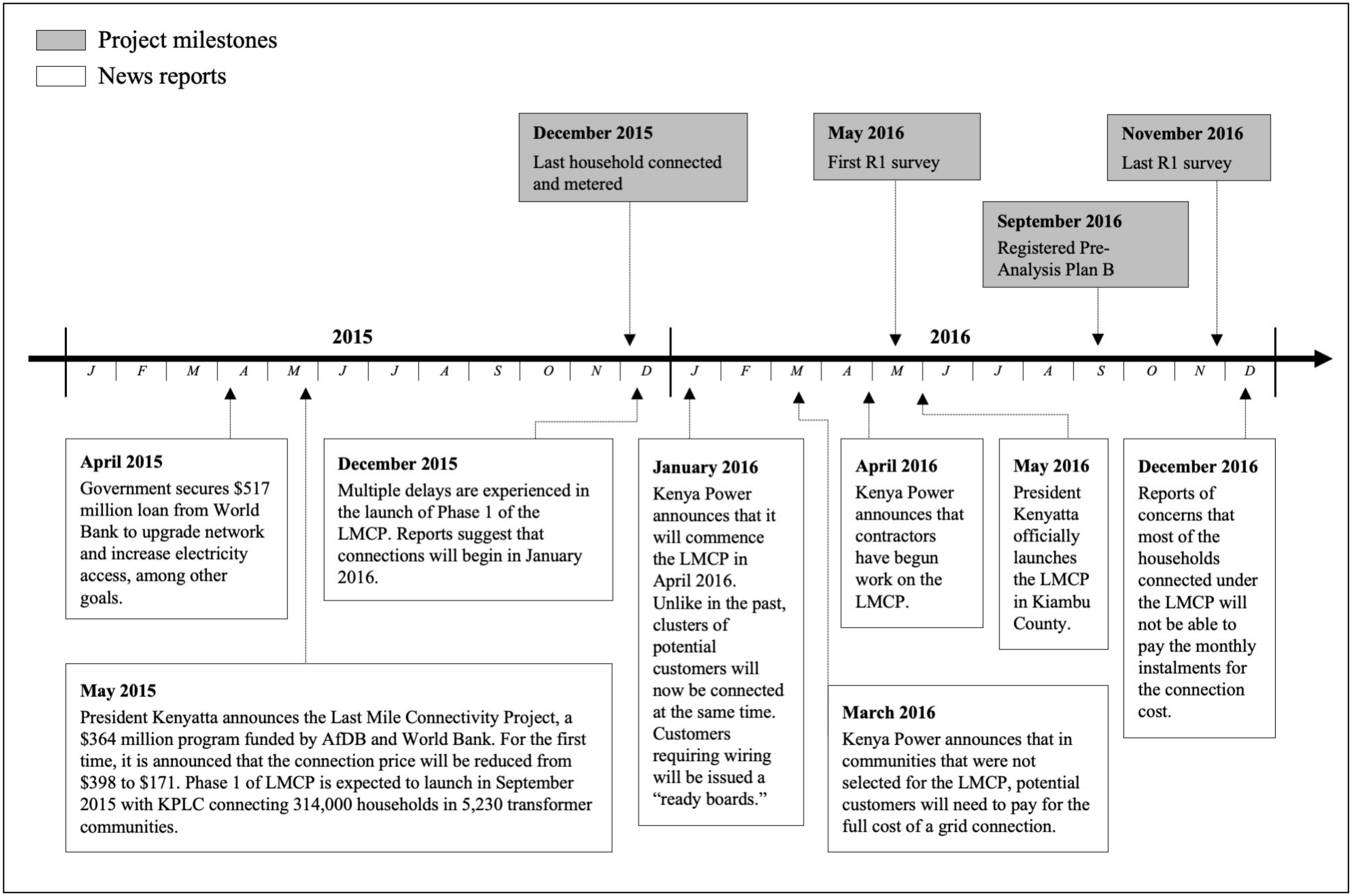
Notes: Treatment households received an opportunity to install a certified household wiring solution in their homes at no additional cost. 88.5 percent of the households connected in the experiment accepted this offer, while 11.5 percent provided their own wiring. Each ready-board, valued at roughly $34 per unit, featured a single light bulb socket, two power outlets, and two miniature circuit breakers. The unit is first mounted onto a wall and the electricity service line is directly connected to the back. The hardware was designed and produced by Power Technics, an electronic supplies manufacturer in Nairobi.

Figure B6—Timeline of project milestones and connection price-related news reports over the period of study



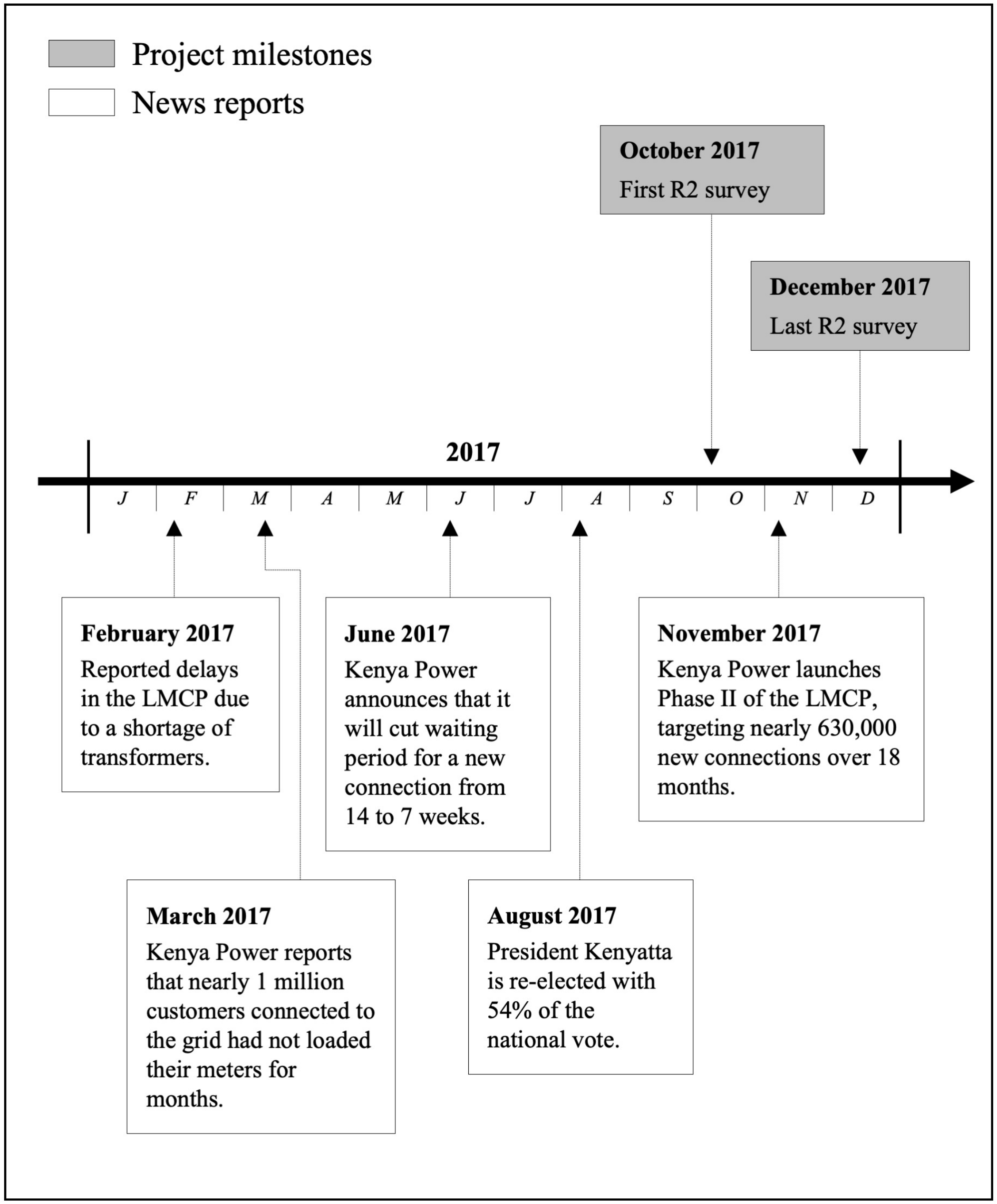
(*Figure continued on next page*)

(*Figure continued from previous page*)



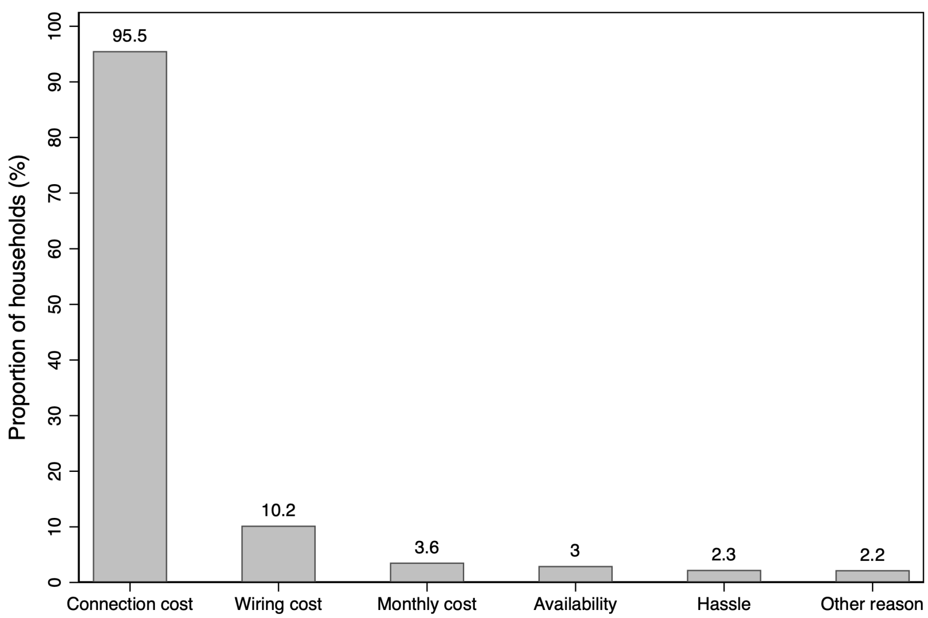
(*Figure continued on next page*)

(*Figure continued from previous page*)



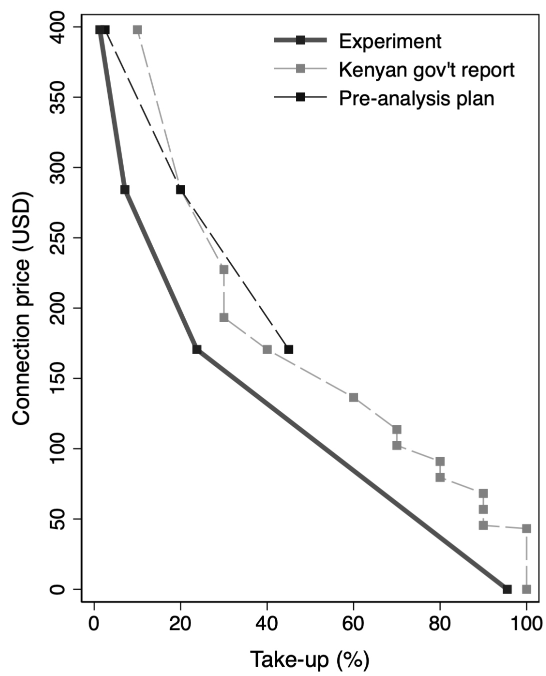
Notes: Sources include *Daily Nation* and *Business Daily*. Note that Pre-Analysis Plan C was registered in March 2018.

Figure B7—Stated reasons why households remain unconnected to electricity at baseline



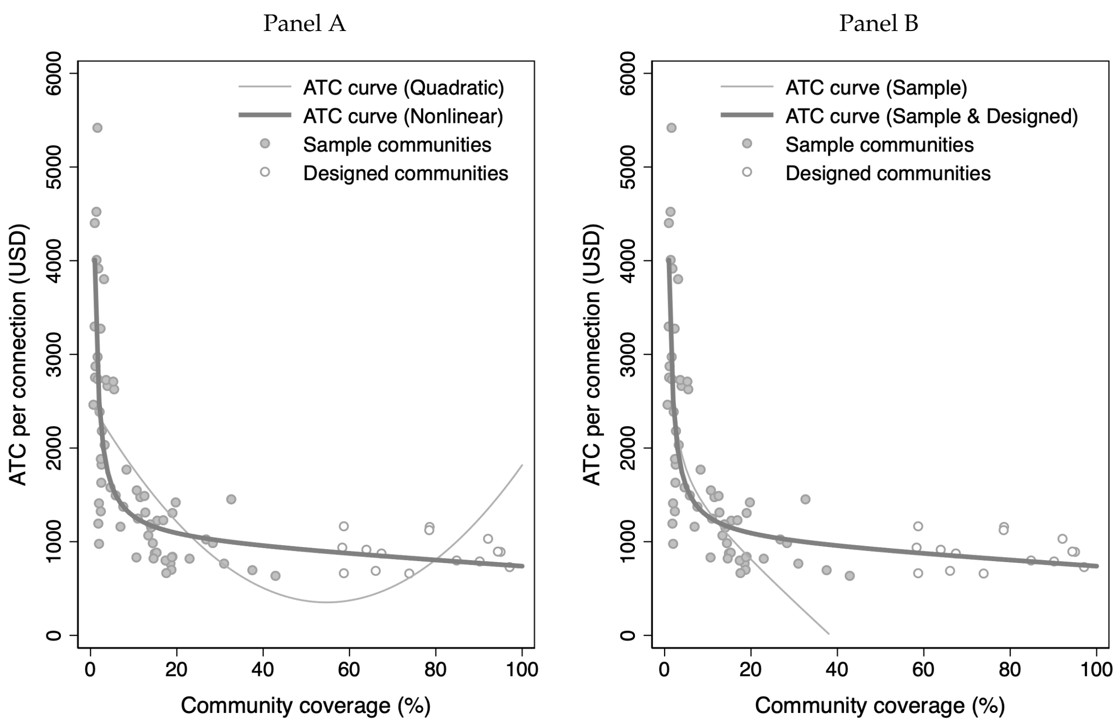
Notes: Based on the responses of 2,289 unconnected households during the baseline survey round.

Figure B8—Experimental evidence on the demand for rural electrification



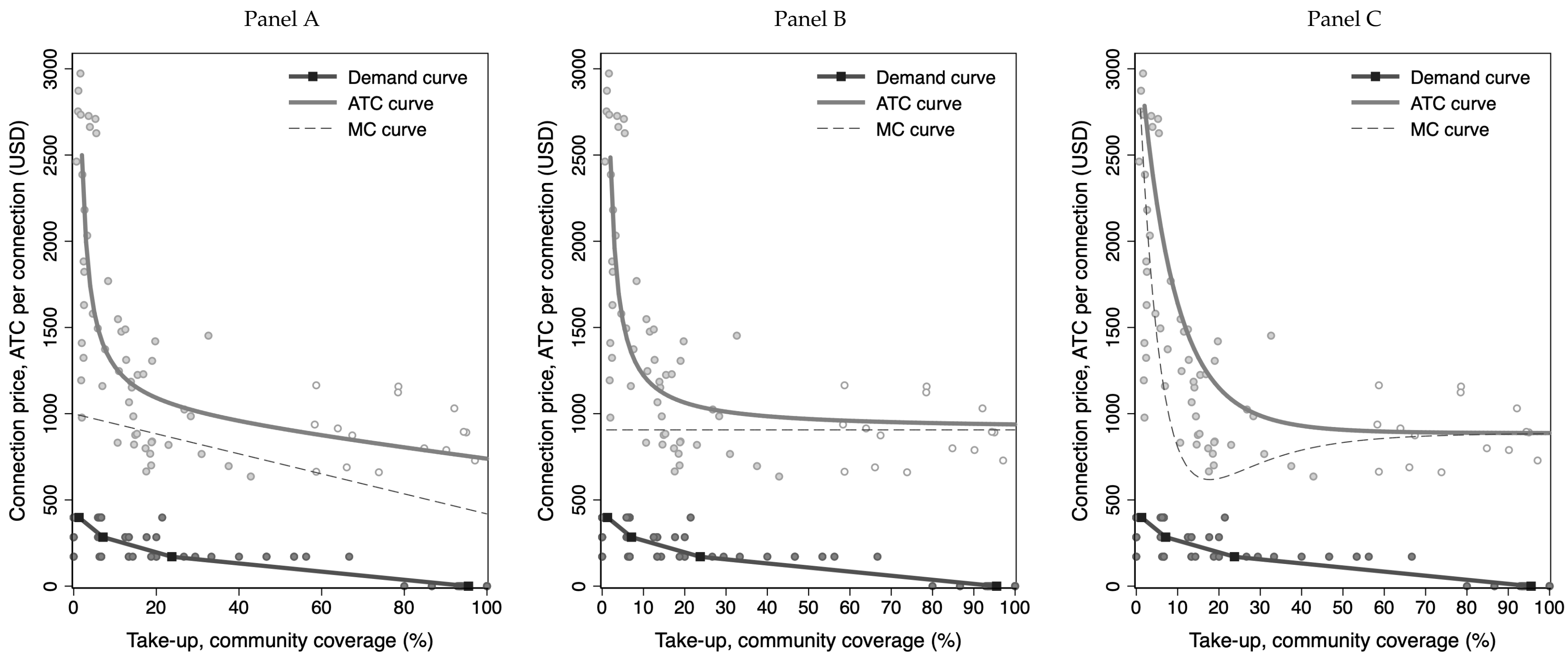
Notes: The experimental results are compared with two sets of initial assumptions based on (i) our pre-analysis plan (see appendix C), and (ii) an internal government report shared with our team in early-2015.

Figure B9A—Experimental evidence on the costs of rural electrification



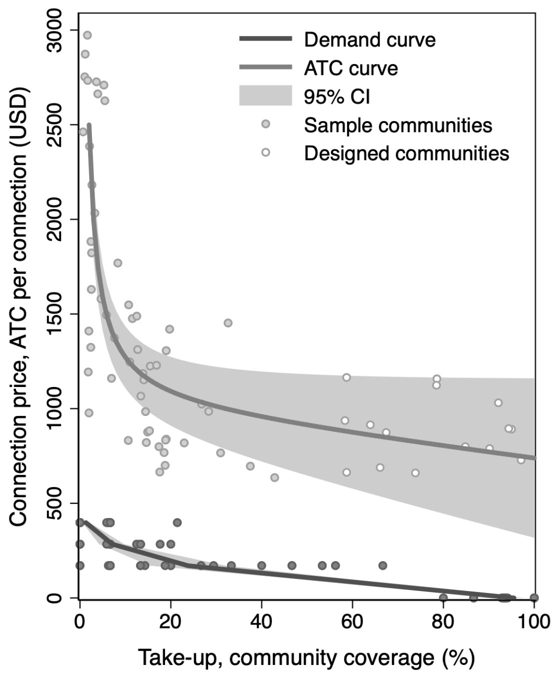
Notes: Each point represents the community-level, budgeted estimate of the average total cost per connection (ATC) at a specific level of community coverage. In panel A, the light-grey curve and the dark-grey curve represent population-weighted ATC curves corresponding to the predicted cost of connecting various population shares, based on the OLS estimation presented in appendix table A1B, column 1, and the nonlinear estimation of $ATC = b_0/M + b_1 + b_2M$, respectively. In panel B, the light-grey curve and the dark-grey curve represent nonlinear ATC curves using the sample communities data (n=62) and the combined sample and designed communities data (N=77), respectively.

Figure B9B—Experimental estimates of a natural monopoly: Alternative functional forms



Notes: Panel A is a reproduction of figure 3, panel A. In panel B, we estimate an ATC curve assuming constant variable costs. In Panel C, we estimate an ATC curve based on an exponential function. In all cases, we plot the population-weighted ATC curve corresponding to the predicted cost of connecting various population shares.

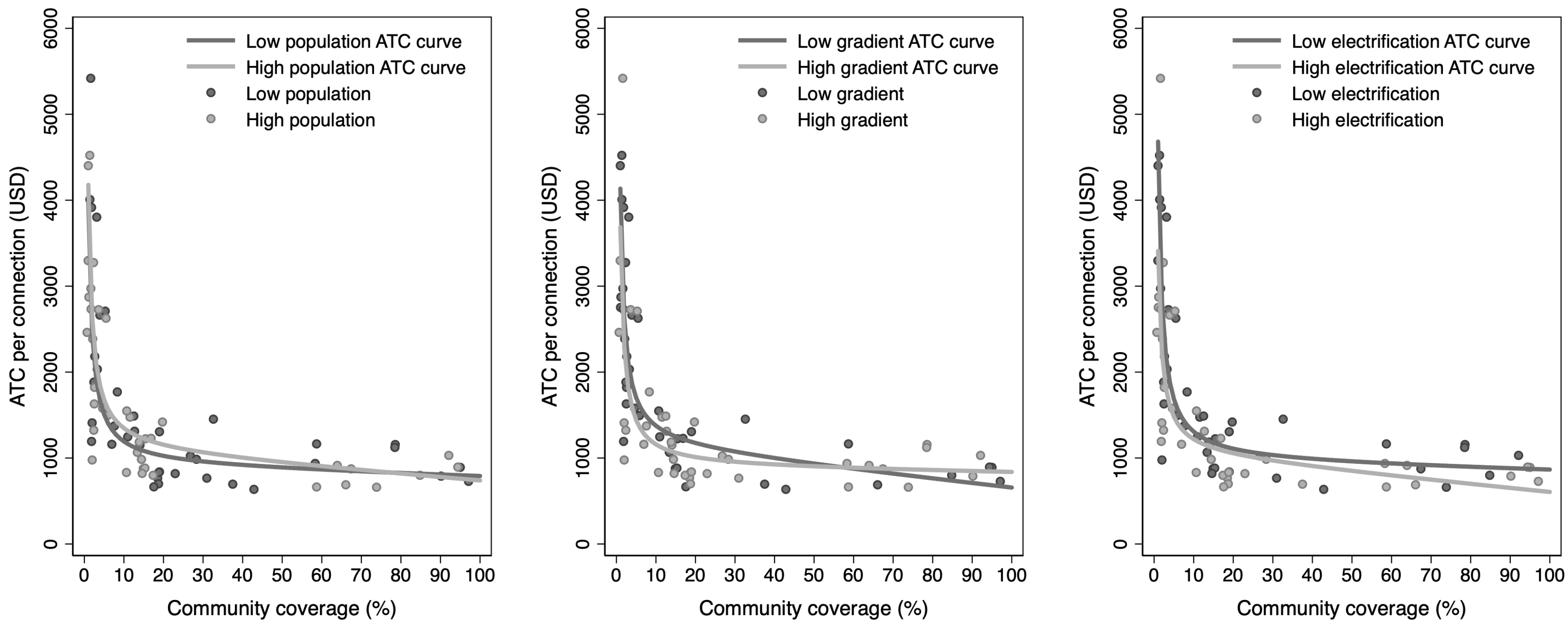
Figure B9C—Experimental estimates of a natural monopoly: Confidence intervals



Notes: The demand curve and ATC curve from figure 3, panel A are plotted with their associated 95 percent confidence intervals. Each point in the demand scatterplot represents a community-level mean (at each price, we show the 95 percent confidence interval around the sample mean). We plot the population-weighted ATC curve and confidence interval corresponding to the predicted cost of connecting various population shares, based on the nonlinear estimation of $ATC = b0/M + b1 + b2M$.

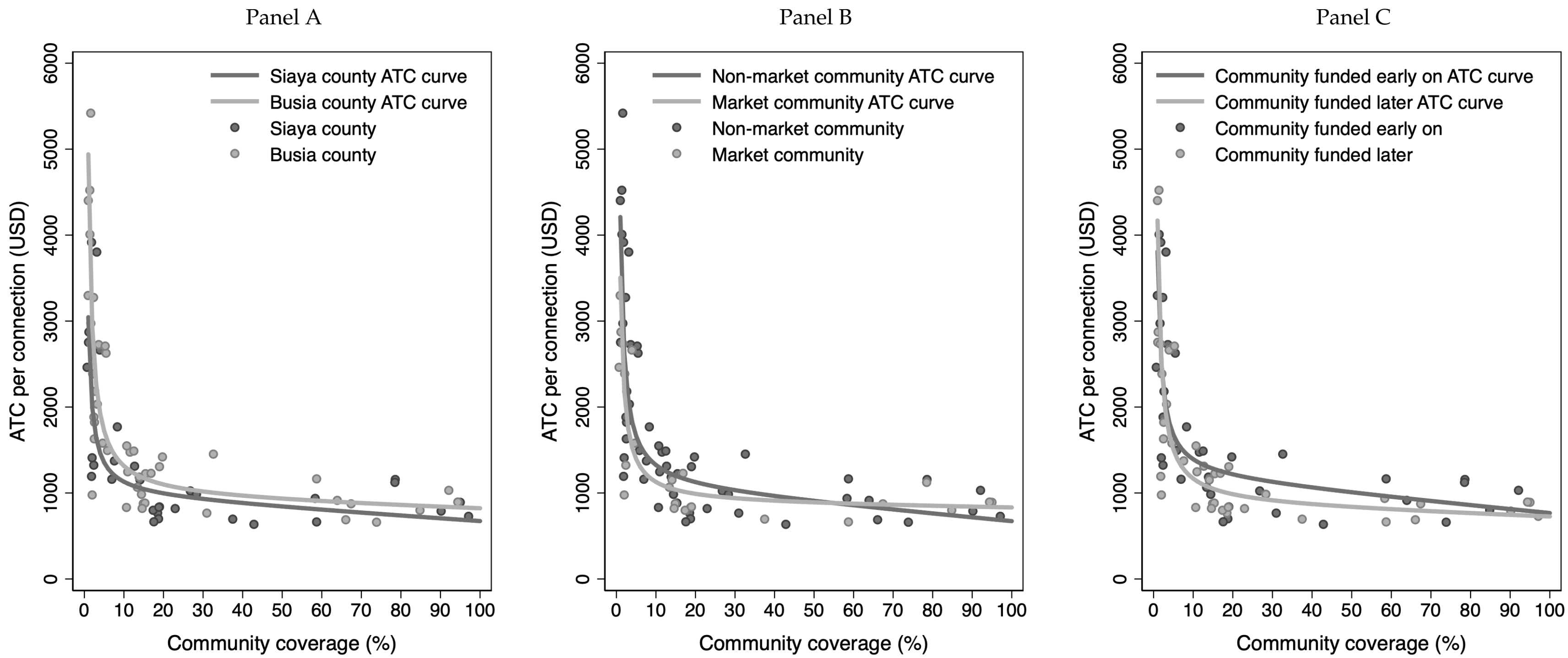
Figure B10A—Comparing the average total cost per connection (ATC) curve for different subsamples

Panel A Panel B Panel C



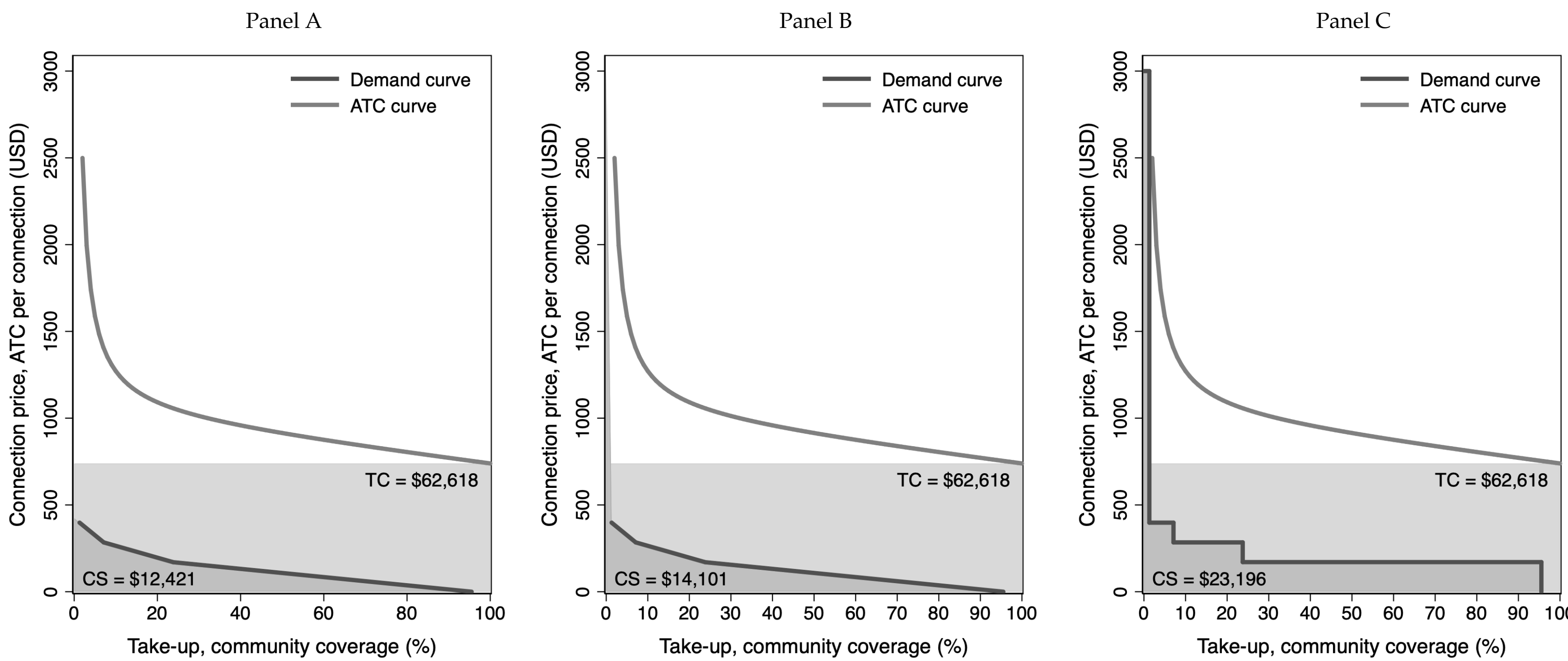
Notes: In panels A, B, and C, we divide the sample into communities with "low" (i.e., below median) and "high" (above median) land gradient, population, and baseline electrification rate, respectively. In panel A, average land gradient ranges from 0.79 to 7.76 degrees with a mean of 2.15 degrees. In each panel, we plot the population-weighted ATC curves for the "low" and "high" subsamples separately.

Figure B10B—Comparing the average total cost per connection (ATC) curve for different subsamples



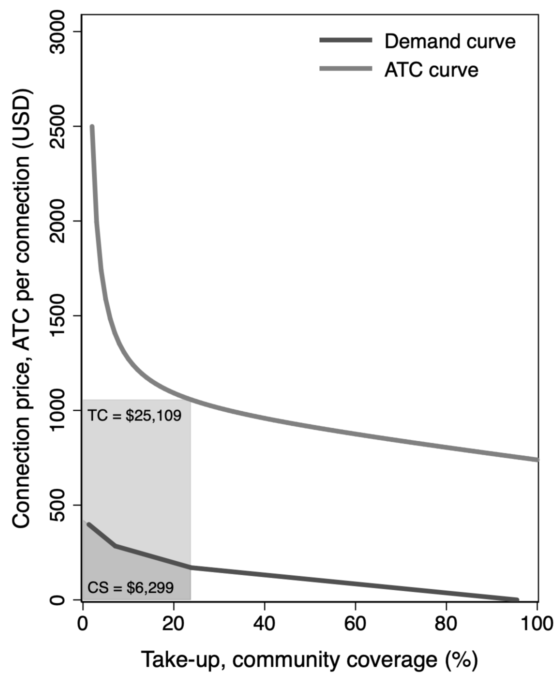
Notes: In panels A, B, and C, we divide the sample according to the county, whether the community is a market center, and whether the community's central transformer was funded (and installed) early on, respectively. We then plot the population-weighted ATC curves for each subsamples separately.

Figure B11—Social surplus implications of rural electrification under various demand curve assumptions



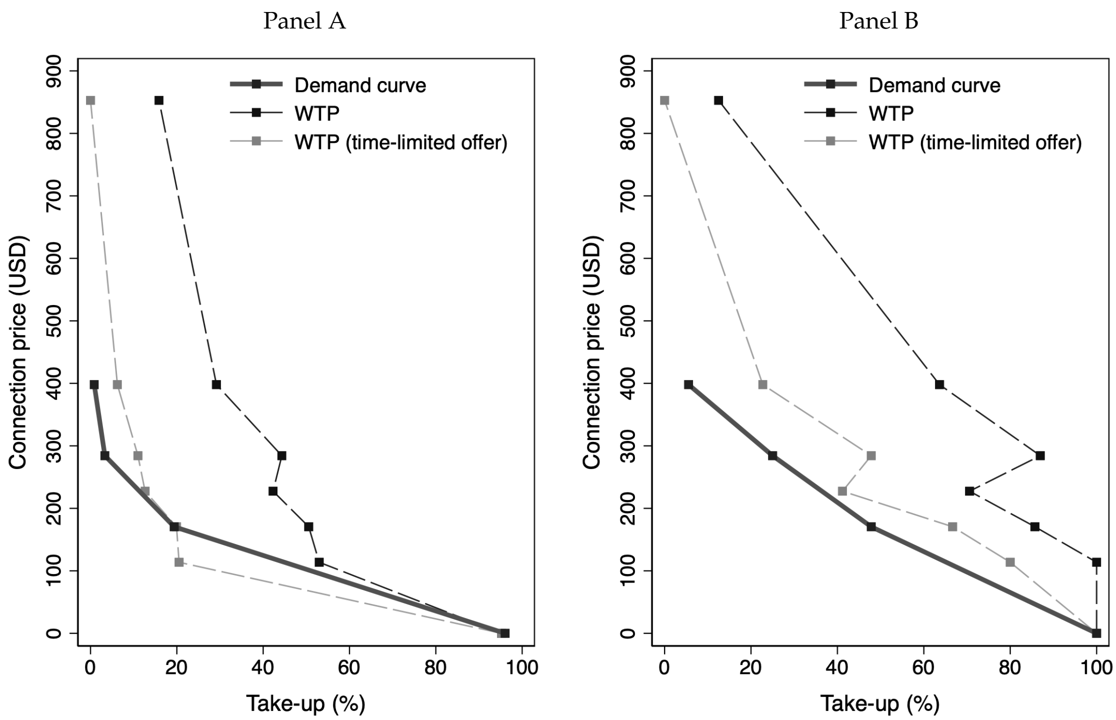
Notes: Panel A is a reproduction of figure 3, Panel B. In this scenario, the implied social surplus loss associated with a mass electrification program is $50,197 per community. In panel B, we estimate the area under the unobserved [0, 1.3] domain by assuming that the demand curve intercepts the vertical axis at $3,000, rather than $424 (as in panel A). In this more conservative case, the implied social surplus loss is $48,516 per community. In order to overturn this result (i.e. costs exceeding the consumer surplus), the intercept would need to be an astronomical $37,594. In panel C, the most conservative case, we assume that demand is a step function. The implied social surplus loss is $39,422 per community. The discounted future social surplus gain needed for consumer surplus to exceed total costs across the three scenarios ranges from $465 (in panel C) to $593 (in panel A) per household.

Figure B12—Social surplus implications of a government program



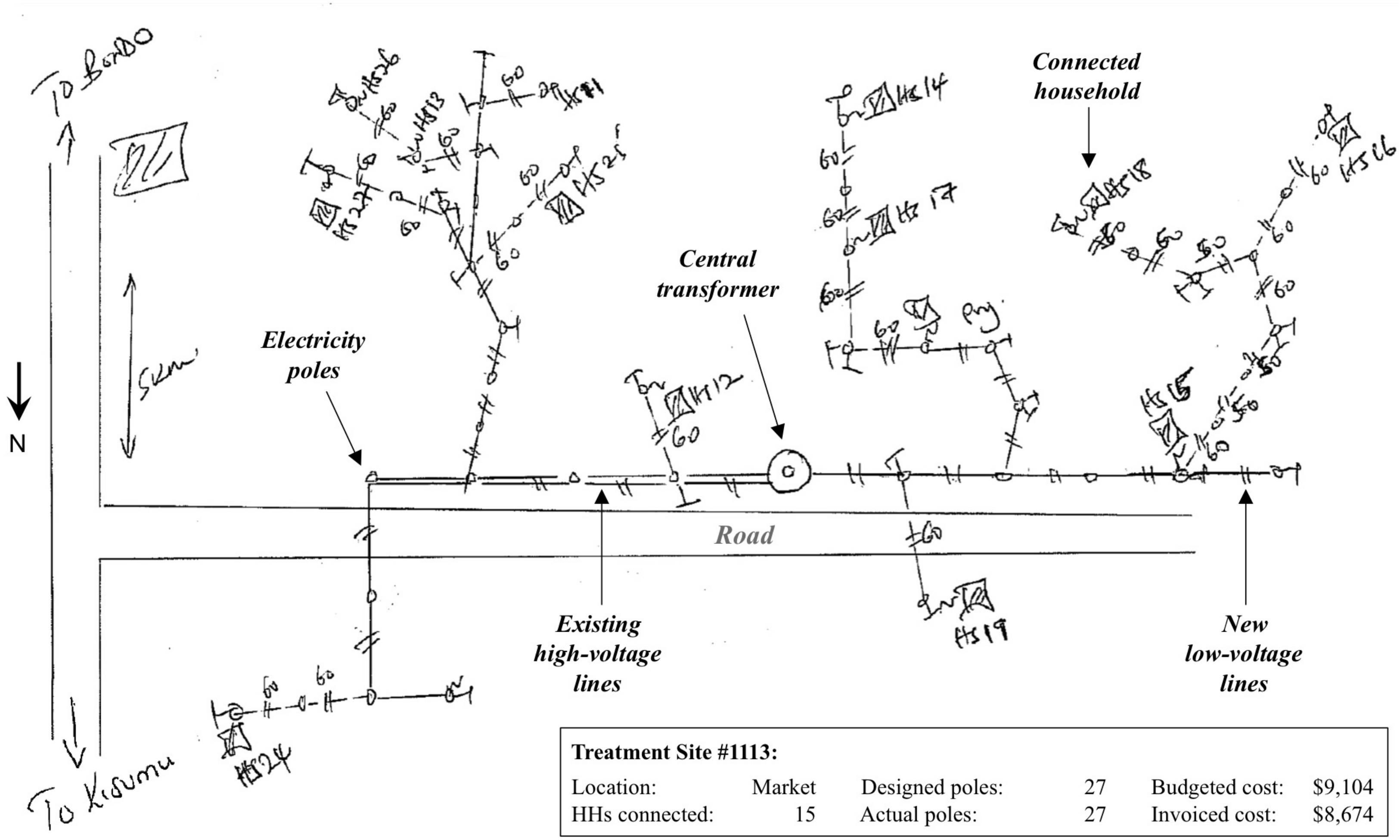
Notes: This figure presents the estimated demand for and costs of a program structured like the planned *Last Mile Connectivity Project*, which offers households a fixed price of $171. In this case, only 23.7 percent of households would accept the price. The results suggest that unless the government is willing to provide additional subsidies, the resulting electrification level would be low and there would be a social surplus loss of $18,809 per community. Discounted average future social surplus gains of $935 would be required per household.

Figure B13—Comparison of demand between households without bank accounts and with low-quality walls (Panel A), and households with bank accounts and high-quality walls (Panel B)



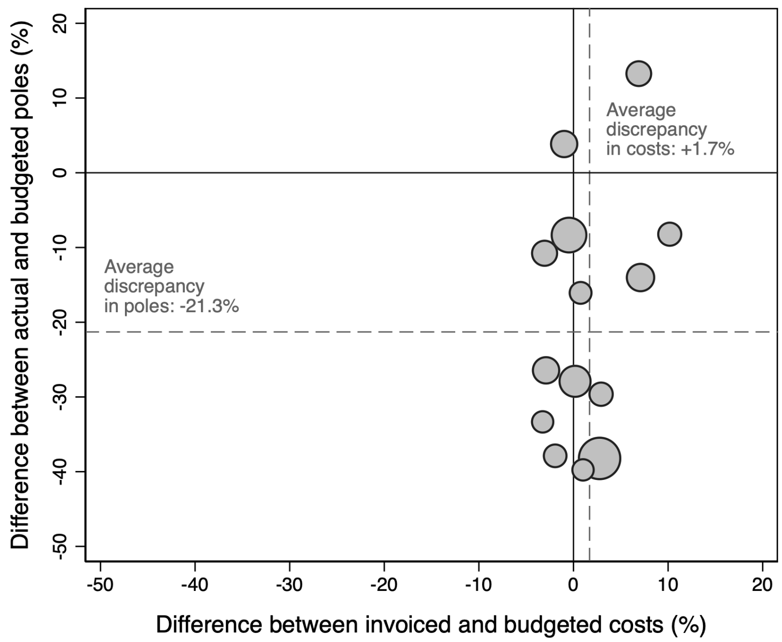
Notes: We plot the experimental results (solid black line) and responses to the contingent valuation questions included in the baseline survey. Households were first asked whether they would accept a hypothetical offer (i.e., randomly assigned price) to connect to the grid (dashed line, black squares). Households were then asked whether they would accept the same hypothetical offer if required to complete the payment in six weeks (dashed line, grey squares). Panel A presents demand curves for households without bank accounts and with low-quality walls. Panel B presents demand curves for households with bank accounts and high-quality walls.

Figure B14—Example of a REA design drawing in a high subsidy treatment community



Notes: After receiving payment, REA designers visited each treatment community to design the local low-voltage network. The designs were then used to estimate the required materials and determine a budgeted estimates of the total construction cost. Materials (e.g. poles, electricity line, service cables) represented 65.9 percent of total installation costs. The community in this example is the same as that shown in appendix figure B2.

Figure B15—Discrepancies in project costs and electrical poles, by contractor



Notes: Each circle represents one of the 14 contractors that participated in the overall project. The size of each circle is proportional to the number of household connections supplied by the contractor (mean=34). The horizontal axis represents the percentage difference between the total invoiced and budgeted cost for each contractor. The vertical axis represents the percentage difference between the actual and designed poles (i.e. materials) for each contractor. The average discrepancies in poles and costs are weighted by the number of connections per contractor and correspond to the values in appendix table B11.

Table B1—Comparison of social and economic indicators for study region and nationwide counties

	Study region	Nationwide county percentiles 25th	50th	75th
Total population	793,125	528,054	724,186	958,791
per square kilometer	375.4	39.5	183.2	332.9
% rural	85.7	71.6	79.5	84.4
% at school	44.7	37.0	42.4	45.2
% in school with secondary education	10.3	9.7	11.0	13.4
Total households	176,630	103,114	154,073	202,291
per square kilometer	83.6	7.9	44.3	78.7
% with high quality roof	59.7	49.2	78.5	88.2
% with high quality floor	27.7	20.6	29.7	40.0
% with high quality walls	32.2	20.3	28.0	41.7
% with piped water	6.3	6.9	14.2	30.6
Total public facilities	644	356	521	813
per capita (000s)	0.81	0.59	0.75	0.98
Electrification rates				
Rural (%)	2.3	1.5	3.1	5.3
Urban (%)	21.8	20.2	27.2	43.2
Public facilities (%)	84.1	79.9	88.1	92.6

Notes: The study region column presents weighted-average and average (where applicable) statistics for Busia and Siaya counties. Specifically, total population, total households, and total public facilities represent averages for Busia and Siaya. We exclude Nairobi and Mombasa, two counties that are entirely urban, from the nationwide county percentile columns. Demographic data is obtained from the 2009 Kenya Population and Housing Census (KPHC). Data on public facilities (defined as market centers, secondary schools, and health clinics) are obtained from the Rural Electrification Authority (REA). High quality roof indicates roofs made of concrete, tiles, or corrugated iron sheets. High quality floor indicates floors made of cement, tiles, or wood. High quality walls indicates walls made of stone, brick, or cement. Rural and urban electrification rates represent the proportion of households that stated that electricity was their main source of lighting during the 2009 census. Based on the 2009 census data, the mean (county-level) electrification rates in rural and urban areas were 4.6 and 32.6 percent, respectively. Nationally, the rural and urban electrification rates were 5.1 and 50.4 percent, respectively, and 22.7 percent overall. An earlier version of this table is presented in Lee et al. (2016).

Table B2—Baseline summary statistics and randomization balance check

		Regression coefficients on subsidy treatment indicators			
	Control	Low	Medium	High	p-value of F-test
	(1)	(2)	(3)	(4)	(5)
Panel A: Household head (respondent)					
Female=1	0.63	0.02	-0.03	-0.02	0.62
	[0.48]	(0.03)	(0.03)	(0.03)	
Age (years)	52.0	-1.1	1.0	1.7	0.28
	[16.3]	(1.2)	(1.1)	(1.4)	
Senior citizen=1	0.27	-0.01	0.00	0.02	0.89
	[0.45]	(0.03)	(0.03)	(0.04)	
Attended secondary school=1	0.14	-0.01	0.03	-0.03	0.29
	[0.34]	(0.02)	(0.03)	(0.03)	
Married=1	0.66	-0.01	0.01	-0.02	0.86
	[0.47]	(0.03)	(0.03)	(0.03)	
Not a farmer=1	0.23	0.00	-0.03	0.00	0.79
	[0.42]	(0.04)	(0.03)	(0.03)	
Employed=1	0.36	0.00	-0.00	0.01	0.98
	[0.48]	(0.03)	(0.03)	(0.03)	
Basic political awareness=1	0.13	-0.05***	-0.01	-0.03	0.04
	[0.33]	(0.02)	(0.02)	(0.02)	
Has bank account=1	0.19	-0.03	0.00	-0.02	0.45
	[0.39]	(0.02)	(0.03)	(0.03)	
Monthly earnings (USD)	16.82	3.94	-2.01	-1.40	0.60
	[53.74]	(4.03)	(3.25)	(3.08)	
Panel B: Household characteristics					
Number of members	5.3	-0.3*	0.1	-0.3	0.07
	[2.7]	(0.1)	(0.2)	(0.2)	
Youth members (age $\leqq$ 18)	3.0	-0.1	0.1	-0.2	0.24
	[2.2]	(0.1)	(0.1)	(0.1)	
High-quality walls=1	0.15	0.05**	0.04	-0.01	0.09
	[0.36]	(0.03)	(0.03)	(0.03)	
Land (acres)	1.9	0.30	0.2	0.1	0.41
	[2.1]	(0.2)	(0.2)	(0.1)	
Distance to transformer (m)	348.6	14.8	9.5	22.1**	0.17
	[140.0]	(9.9)	(12.2)	(10.6)	
Monthly (non-charcoal) energy (USD)	5.55	-0.23	0.50*	-0.43	0.02
	[5.20]	(0.27)	(0.27)	(0.28)	

(*Table continued on next page*)

(*Table continued from previous page*)

	Control	Low	Medium	High	p-value of F-test
		Regression coefficients on subsidy treatment indicators			
	(1)	(2)	(3)	(4)	(5)
Panel C: Household assets					
Bednets	2.3	0.0	0.1	0.0	0.89
	[1.5]	(0.1)	(0.1)	(0.1)	
Bicycles	0.7	0.0	0.1	0.0	0.35
	[0.7]	(0.0)	(0.1)	(0.1)	
Sofa pieces	5.9	0.0	0.5	0.0	0.66
	[5.2]	(0.4)	(0.4)	(0.4)	
Chickens	7.0	0.4	-0.4	-0.2	0.74
	[8.7]	(0.7)	(0.6)	(0.7)	
Cattle	1.7	0.1	0.2	0.2	0.51
	[2.3]	(0.2)	(0.2)	(0.2)	
Owns radio=1	0.34	-0.02	0.05	0.00	0.41
	[0.48]	(0.03)	(0.03)	(0.04)	
Owns television=1	0.16	0.00	0.00	-0.05**	0.13
	[0.37]	(0.02)	(0.03)	(0.02)	
Panel D: Community characteristics					
Community electrification rate (%)	5.3	1.6	0.0	-0.1	0.67
	[4.6]	(1.3)	(1.0)	(0.9)	
Community population	534.7	42.1	26.4	9.8	0.79
	[219.0]	(45.0)	(41.7)	(39.1)	

Notes: Column 1 reports mean values for the control group, with standard deviations in brackets. Columns 2 to 4 report the coefficients from separate regressions in which a dependent variable is regressed on the full set of treatment indicators and stratification variables (i.e., county, market status, and whether the transformer was funded and installed early on, between 2008 and 2010). Standard errors are in parantheses. Column 5 reports the p-values of F-tests of whether the treatment coefficients are jointly equal to zero. Robust standard errors clustered at the community level. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$. Sample sizes range from 2,275 to 2,289 depending on missing values except in the specification with age as the dependent variable where the sample size is 2,205. Monthly earnings (USD) includes the respondent's profits from businesses and self-employment, salary and benefits from employment, and agricultural sales for the entire household. An overall F-test in an SUR specification across the 25 regressions yields a p-value on the F-statistic of 0.64; we cannot reject the hypothesis of baseline equality across all of the treatment arms and control groups. Only 11 of the variables listed in this table were pre-specified. An F-test across these variables yields a p-value of 0.07; we again cannot reject the hypothesis of baseline equality at the standard 95 percent confidence level.

Table B3—Characteristics of households taking-up electricity by treatment arm

	High subsidy Price: \$0 (1)	Medium subsidy Price: *\$171* (2)	Low subsidy Price: *\$284* (3)	Control Price: \$398 (4)
Panel A: Respondent characteristics				
Female (%)	61.7	58.9	59.3	60.0
Age (years)	53.7	52.8	50.6	51.6
Senior citizen (%)	28.9	24.4	25.9	28.6
Attended secondary school (%)	9.9	27.8***	33.3***	26.7**
Married (%)	64.2	74.4*	70.4	66.7
Not a farmer (%)	22.3	28.9	29.6	28.6
Employed (%)	36.4	45.6	55.6**	66.7**
Basic political awareness (%)	9.6	16.7*	14.8	6.7
Has bank account (%)	17.1	31.1***	40.7***	35.7*
Monthly earnings (USD)	14.4	26.0*	77.9***	45.8**
Panel B: Household characteristics				
Number of members	5.0	6.2***	6.2**	5.8
Youth members (age $\leqq$ 18)	2.8	3.5***	3.9**	3.3
High-quality walls (%)	13.0	25.6***	51.9***	33.3**
Land (acres)	1.9	2.2	2.6	2.1
Distance to transformer (m)	369.7	357.4	369.1	360.7
Monthly (non-charcoal) energy (USD)	5.2	7.6***	8.2***	5.9
Panel C: Household assets				
Bednets	2.3	2.8***	3.4***	2.5
Sofa pieces	5.9	9.0***	9.4***	8.9**
Chickens	6.9	9.1**	10.3*	6.4
Owns radio (%)	33.6	47.8**	48.2	53.3
Owns television (%)	10.7	27.8***	48.2***	40.0***
Take-up of electricity connections	363	90	27	15

Notes: Columns 1, 2, and 3 report sample means for unconnected households that chose to take-up a subsidized electricity connection. Column 4 reports sample means for control group households that chose to connect on their own. Basic political awareness indicator captures whether the household head was able to correctly identify the heads of state of Tanzania, Uganda, and the United States. Monthly earnings (USD) includes the respondent's profits from businesses and self-employment, salary and benefits from employment, and agricultural sales for the entire household. The asterisks in columns 2, 3, and 4 indicate statistically significant differences compared to column 1: * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table B4A—Impact of connection subsidy on take-up: Interactions with community-level variables

		Interacted variable			
		Busia county	Transformer funded early on	Market center	Baseline population
	(1)	(2)	(3)	(4)	(5)
T1: Low subsidy—29% discount	5.9***	2.7	5.0***	6.3***	2.3
	(1.5)	(1.7)	(1.9)	(1.7)	(4.0)
T2: Medium subsidy—57% discount	22.9***	20.9***	26.8***	23.5***	18.5*
	(4.0)	(5.8)	(6.2)	(4.8)	(10.3)
T3: High subsidy—100% discount	95.0***	95.2***	93.7***	94.9***	100.1***
	(1.3)	(1.7)	(1.7)	(1.6)	(4.5)
Interacted variable		0.2	0.2	0.9	-0.0
		(0.9)	(0.8)	(1.0)	(0.0)
T1 × interacted variable		5.6**	2.1	-1.6	0.0
		(2.7)	(3.1)	(3.3)	(0.0)
T2 × interacted variable		3.5	-8.2	-2.7	0.0
		(8.0)	(7.9)	(9.0)	(0.0)
T3 × interacted variable		-0.4	2.7	0.2	-0.0
		(2.6)	(2.5)	(2.4)	(0.0)
Take-up in control group	1.3	1.3	1.3	1.3	1.3
Observations	2,176	2,176	2,176	2,176	2,176
R-squared	0.69	0.69	0.69	0.69	0.69

Notes: The dependent variable is an indicator variable (multiplied by 100) for household take-up. The mean of the dependent variable is 21.6. Robust standard errors clustered at the community level in parentheses. All specfications include the pre-specified household and community covariates. Household covariates include the age of the household head, indicators for whether the household respondent attended secondary school, is a senior citizen, is not primarily a farmer, is employed, and has a bank account, an indicator for whether the household has high-quality walls, and the number of chickens (a measure of assets) owned by the household. Community covariates include indicators for the county, market status, whether the transformer was funded and installed early on (between 2008 and 2010), community electrification rate at baseline, and community population. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$. The number of observations is somewhat smaller than the total number of households in our sample (2,289) due to missing data. The coefficients do not change appreciably when the households with missing data are included in the specification in column 1.

Table B4B—Impact of connection subsidy on take-up: Interactions with household-level variables

	Interacted variable		
	Household size	Age of household head	Senior household head
	(1)	(2)	(3)
T1: Low subsidy—29% discount	0.6	5.5	5.5***
	(2.7)	(5.0)	(1.7)
T2: Medium subsidy—57% discount	9.8*	26.2***	23.7***
	(5.7)	(7.1)	(4.2)
T3: High subsidy—100% discount	94.2***	95.2***	95.5***
	(2.7)	(3.5)	(1.2)
Interacted variable	0.0	0.0	1.2
	(0.2)	(0.0)	(1.3)
T1 × interacted variable	1.0*	0.0	1.7
	(0.5)	(0.1)	(4.3)
T2 × interacted variable	2.4***	-0.1	-3.1
	(0.9)	(0.1)	(3.6)
T3 × interacted variable	0.1	-0.0	-2.0
	(0.4)	(0.1)	(2.3)
Take-up in control group	1.3	1.3	1.3
Observations	2,176	2,176	2,176
R-squared	0.69	0.69	0.69

Notes: The dependent variable is an indicator variable (multiplied by 100) for household take-up. The mean of the dependent variable is 21.6. Robust standard errors clustered at the community level in parentheses. All specfications include the pre-specified household and community covariates. Household covariates include the age of the household head, indicators for whether the household respondent attended secondary school, is a senior citizen, is not primarily a farmer, is employed, and has a bank account, an indicator for whether the household has high-quality walls, and the number of chickens (a measure of assets) owned by the household. Community covariates include indicators for the county, market status, whether the transformer was funded and installed early on (between 2008 and 2010), community electrification rate at baseline, and community population. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table B4C—Impact of connection subsidy on take-up: Interactions with household-level variables

	Interacted variable		
	Number of chickens	Has bank account	Not a farmer
	(1)	(2)	(3)
T1: Low subsidy—29% discount	4.7***	4.5***	5.4***
	(1.4)	(1.4)	(1.6)
T2: Medium subsidy—57% discount	17.2***	20.3***	20.1***
	(3.8)	(4.1)	(4.6)
T3: High subsidy—100% discount	93.8***	94.9***	94.9***
	(1.8)	(1.4)	(1.4)
Interacted variable	-0.1*	1.1	-0.7
	(0.0)	(1.2)	(0.9)
T1 × interacted variable	0.2	8.4	2.4
	(0.1)	(5.9)	(3.6)
T2 × interacted variable	0.8***	13.5*	13.5*
	(0.3)	(7.3)	(7.7)
T3 × interacted variable	0.2	-0.0	0.3
	(0.1)	(2.5)	(2.4)
Take-up in control group	1.3	1.3	1.3
Observations	2,176	2,176	2,176
R-squared	0.70	0.69	0.69

Notes: The dependent variable is an indicator variable (multiplied by 100) for household take-up. The mean of the dependent variable is 21.6. Robust standard errors clustered at the community level in parentheses. All specfications include the pre-specified household and community covariates. Household covariates include the age of the household head, indicators for whether the household respondent attended secondary school, is a senior citizen, is not primarily a farmer, is employed, and has a bank account, an indicator for whether the household has high-quality walls, and the number of chickens (a measure of assets) owned by the household. Community covariates include indicators for the county, market status, whether the transformer was funded and installed early on (between 2008 and 2010), community electrification rate at baseline, and community population. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table B4D—Impact of connection subsidy on take-up: Full list of controls

	(1)	(2)	(3)
Control (intercept)	1.3***	-9.5**	-10.6**
	(0.4)	(3.9)	(4.7)
T1: Low subsidy—29% discount	5.8***	5.9***	6.2***
	(1.4)	(1.5)	(1.5)
T2: Medium subsidy—57% discount	22.4***	22.9***	22.7***
	(4.0)	(4.0)	(4.0)
T3: High subsidy—100% discount	94.2***	95.0***	95.1***
	(1.2)	(1.3)	(1.3)
Female=1			0.8
			(1.3)
Age (years)		0.0	0.0
		(0.0)	(0.0)
Senior citizen=1		0.5	1.1
		(1.4)	(1.5)
Attended secondary school=1		3.8**	3.3**
		(1.7)	(1.7)
Married=1			-1.5
			(1.2)
Not a farmer=1		1.9	1.8
		(1.6)	(1.5)
Employed=1		1.1	-0.1
		(1.3)	(1.3)
Basic political awareness=1			-1.4
			(1.5)
Has bank account=1		2.6	1.5
		(1.7)	(1.6)
Monthly earnings (USD)			0.0
			(0.0)
Number of members		0.6***	0.5
		(0.2)	(0.4)
Youth members (age $\leq$ 18)			-0.1
			(0.5)
High-quality walls=1		3.5	0.9
		(2.1)	(2.1)

(*Table continued on next page*)

(*Table continued from previous page*)

	(1)	(2)	(3)
Land (acres)			-0.2
			(0.2)
Distance to transformer (100 meters)			-0.2
			(0.3)
Monthly (non-charcoal) energy (USD)			0.2
			(0.1)
Number of bednets			0.4
			(0.5)
Number of bicycles			1.7*
			(0.9)
Number of sofa pieces			0.3**
			(0.1)
Number of chickens		0.1**	0.1
		(0.1)	(0.1)
Number of cattle			-0.1
			(0.3)
Owns radio=1			-0.6
			(1.0)
Owns television=1			2.7*
			(1.6)
Electrification rate (%)		0.1	0.1
		(0.2)	(0.2)
Community population		0.0	0.0
		(0.0)	(0.0)
Busia=1		1.7	2.0
		(1.5)	(1.5)
Funded and installed early on=1		-0.5	-0.8
		(1.6)	(1.6)
Market status=1		0.2	0.5
		(1.6)	(1.7)
Observations	2,289	2,176	2,162
R-squared	0.68	0.69	0.70

Notes: The dependent variable is an indicator variable (multiplied by 100) for household take-up, with a mean of 21.6. Robust standard errors clustered at the community level in parentheses. Column 2 includes pre-specified household and community controls. Column 3 includes both pre-specified controls and additional characteristics listed in appendix table B2. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table B4E—Impact of grid connection price on take-up

	(1)	(2)
Price	-0.5***	-0.5***
	(0.0)	(0.0)
Price^2 $\times$ 1000	0.7***	0.7***
	(0.1)	(0.1)
Age (years)		0.0
		(0.0)
Senior citizen=1		0.6
		(1.4)
Attended secondary school=1		3.6**
		(1.7)
Not a farmer=1		1.9
		(1.6)
Has bank account=1		2.5
		(1.7)
Employed=1		1.1
		(1.3)
Number of members		0.6***
		(0.2)
High-quality walls=1		3.6*
		(2.2)
Number of chickens		0.1**
		(0.1)
Busia=1		1.8
		(1.5)
Funded early on=1		-0.5
		(1.6)
Market status=1		0.3
		(1.6)
Electrification rate (%)		0.2
		(0.2)
Community population		0.0
		(0.0)
Observations	2,289	2,176
R-squared	0.68	0.69

Notes: The dependent variable is an indicator variable (multiplied by 100) for household take-up, with a mean of 21.6. Robust standard errors clustered at the community level in parentheses. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table B5—Actual versus fitted total cost and ATC values (at various coverage levels)

	Mean coverage levels (sample communities)			Coverage benchmarks (sample & designed communities)			
	2.1%	4.8%	17.1%	25%	50%	75%	100%
	T1: Low	*T2: Medium*	*T3: High*				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: REA contractor invoices							
ATC	2,828	2,045	1,000	–	–	–	–
Total cost	4,699	6,419	14,591	–	–	–	–
Panel B: Nonlinear estimates in figure 3 (population-weighted)							
ATC	2,427	1,613	1,126	1,047	915	822	739
Total cost	4,317	6,557	16,307	22,177	38,731	52,211	62,618
Panel C: OLS estimates in figure B9A, panel A and table A1A, column 1 (population-weighted)							
ATC	2,333	2,134	1,361	979	363	645	1,826
Total cost	4,149	8,675	19,708	20,730	15,353	40,970	154,683

Notes: Columns 1 to 3 report total cost (corresponding to each coverage level) and the average total cost per connection (ATC) based on the mean coverage levels achieved in the experiment. Columns 4 to 7 report fitted total cost and ATC at various benchmarks, based on nonlinear (panel B) and OLS (panel C) regressions using data from both the sample and designed communities.

Table B6A—Round 1 treatment effects on key outcomes

	Control	ITT	TOT	FDR q-val
	(1)	(2)	(3)	(4)
Panel A: Primary energy outcomes				
A1. Grid connected (%)	5.6	89.7***	–	–
	[23.0]	(1.4)		
A2. Monthly electricity spending (USD)	0.14	1.93***	2.14***	–
	[0.92]	(0.13)	(0.14)	
Panel B: Additional energy outcomes				
B1. Electricity as main lighting source (%)	5.2	80.2***	89.2***	0.001
	[22.2]	(2.3)	(2.1)	
B2. Number of appliance types owned	1.8	0.3**	0.3**	0.018
	[1.3]	(0.1)	(0.1)	
B3. Owns mobile phone (%)	84.3	-3.3*	-3.0	0.198
	[36.4]	(2.0)	(2.2)	
B4. Owns radio (%)	54.2	3.5	4.2	0.198
	[49.8]	(2.8)	(3.1)	
B5. Owns television (%)	17.9	9.7***	10.9***	0.004
	[38.3]	(2.9)	(3.4)	
B6. Owns iron (%)	4.1	2.2**	2.4**	0.066
	[19.9]	(1.1)	(1.2)	
B7. Monthly kerosene spending (USD)	2.81	-1.15***	-1.20***	0.001
	[2.86]	(0.16)	(0.17)	
B8. Monthly total energy spending (USD)	11.66	-0.35	-0.40	0.832
	[28.47]	(1.69)	(1.86)	
B9. Solar home system as main lighting source (%)	11.8	-11.2***	-12.8***	0.001
	[32.3]	(1.4)	(1.4)	
Panel C: Primary economic outcomes				
C1. Household employed or own business (%)	36.8	5.1	4.3	0.427
	[38.8]	(3.3)	(3.6)	
C3. Total hours worked last week	47.0	-2.0	-2.5*	0.333
	[24.7]	(1.3)	(1.4)	
C4. Total asset value (USD)	914	140	148	0.427
	[961]	(121)	(137)	
C5. Per capita consumption of major items (USD)	133	-7	-9	0.427
	[142]	(8)	(9)	

(*Table continued on next page*)

(*Table continued from previous page*)

	Control	ITT	TOT	FDR q-val
	(1)	(2)	(3)	(4)
Panel D: Primary non-economic outcomes				
D1. Recent health symptoms index	0	-0.03	-0.04	0.613
	[1]	(0.06)	(0.07)	
D2. Normalized life satisfaction	0	0.11*	0.12*	0.333
	[1]	(0.06)	(0.07)	
D3. Avg. student test Z-score	0	-0.09	-0.13	0.427
	[1]	(0.09)	(0.10)	
D5. Political and social awareness index	0	-0.03	-0.01	0.861
	[1]	(0.05)	(0.05)	
Panel E: Mean treatment effects on grouped outcomes				
E1. Economic Index (C outcomes)	0	0.06	0.04	–
	[1]	(0.08)	(0.09)	
E2. Non-Economic Index (D outcomes)	0	-0.03	-0.04	–
	[1]	(0.07)	(0.07)	

Notes: Column 1 reports mean values in the control group, with standard deviations in brackets. Column 2 reports coefficients from separate ITT regressions in which the dependent variable (e.g., A1) is regressed on the high subsidy treatment indicator. The low and medium subsidy groups are excluded from these regressions. Sample sizes range from 1,454 to 1,461 for these regressions, except for the D3 regression, in which the sample size is 941. Column 3 reports coefficients from separate TOT (IV) regressions in which household electrification status is instrumented with the three subsidy treatment indicators. Sample sizes range from 2,171 to 2,180 for these regressions, except for the D3 regression, in which the sample size is 1,411. All specifications include pre-specified household, student, and community covariates. Column 4 reports the FDR-adjusted q-values associated with the coefficient estimates in column 3. FDR-adjusted q-values are computed for each outcome within the additional energy outcomes group (panel B), and for each outcome within the primary outcomes group (panels C and D combined). In panel E, we report mean treatment effects on outcomes grouped into an economic and non-economic index. These groupings were not pre-specified. Robust standard errors clustered at the community level in parentheses. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table B6B—Round 2 treatment effects on key outcomes

	Control	ITT	TOT	FDR q-val
	(1)	(2)	(3)	(4)
Panel A: Primary energy outcomes				
A1. Grid connected (%)	18.9	75.8***	–	–
	[39.2]	(2.7)		
A2. Monthly electricity spending (USD)	0.54	1.66***	2.20***	–
	[1.67]	(0.16)	(0.19)	
Panel B: Additional energy outcomes				
B1. Electricity as main lighting source (%)	16.2	63.8***	84.0***	0.001
	[36.8]	(2.7)	(2.9)	
B2. Number of appliance types owned	2.1	0.4***	0.5***	0.001
	[1.5]	(0.1)	(0.1)	
B3. Owns mobile phone (%)	86.1	-1.4	-1.3	0.675
	[34.6]	(1.9)	(2.4)	
B4. Owns radio (%)	61.0	5.9**	11.4***	0.002
	[48.8]	(2.6)	(3.4)	
B5. Owns television (%)	24.7	8.9***	12.6***	0.005
	[43.2]	(3.2)	(4.2)	
B6. Owns iron (%)	6.3	3.7**	5.4**	0.013
	[24.2]	(1.7)	(2.1)	
B7. Monthly kerosene spending (USD)	2.47	-0.63***	-0.74***	0.001
	[2.63]	(0.14)	(0.18)	
B8. Monthly total energy spending (USD)	9.98	-0.48	-0.01	0.987
	[11.58]	(0.63)	(0.77)	
B9. Solar home system as main lighting source (%)	16.5	-14.9***	-19.8***	0.001
	[37.2]	(1.6)	(1.7)	
Panel C: Primary economic outcomes				
C1. Household employed or own business (%)	35.1	0.4	-0.3	0.921
	[38.1]	(2.5)	(3.2)	
C2. Per capita monthly household earnings (USD)	12	-1	-2	0.751
	[42]	(2)	(2)	
C3. Total hours worked last week	53.7	-3.4**	-4.9**	0.092
	[23.6]	(1.6)	(2.1)	
C4. Total asset value (USD)	1,568	61	79	0.540
	[1,154]	(56)	(74)	
C5. Per capita consumption of major items (USD)	238	2	2	0.921
	[209]	(11)	(14)	

(*Table continued on next page*)

(*Table continued from previous page*)

	Control	ITT	TOT	FDR q-val
	(1)	(2)	(3)	(4)
Panel D: Primary non-economic outcomes				
D1. Recent health symptoms index	0	-0.02	-0.01	0.921
	[1]	(0.07)	(0.1)	
D2. Normalized life satisfaction	0	0.22***	0.29***	0.016
	[1]	(0.07)	(0.09)	
D4. Avg. student KCPE test Z-score	0	-0.12	-0.17	0.540
	[1]	(0.13)	(0.17)	
D6. Perceptions of security index	0	0.08	0.13*	0.248
	[1]	(0.06)	(0.08)	
Panel E: Mean treatment effects on grouped outcomes				
E1. Economic Index (C outcomes)	0	-0.03	-0.06	–
	[1]	(0.06)	(0.08)	
E2. Non-Economic Index (D outcomes)	0	0.06	0.07	–
	[1]	(0.06)	(0.07)	

Notes: Column 1 reports mean values in the control group, with standard deviations in brackets. Column 2 reports coefficients from separate ITT regressions in which the dependent variable (e.g., A1) is regressed on the high subsidy treatment indicator. The low and medium subsidy groups are excluded from these regressions. Sample sizes range from 1,419 to 1,433 for these regressions, except for the D4 regression, in which the sample size is 417. Column 3 reports coefficients from separate TOT (IV) regressions in which household electrification status is instrumented with the three subsidy treatment indicators. Sample sizes range from 2,094 to 2,115 for these regressions, except for the D4 regression, in which the sample size is 644. All specifications include pre-specified household, student, and community covariates. Column 4 reports the FDR-adjusted q-values associated with the coefficient estimates in column 3. FDR-adjusted q-values are computed for each outcome within the additional energy outcomes group (panel B), and for each outcome within the primary outcomes group (panels C and D combined). In panel E, we report mean treatment effects on outcomes grouped into an economic and non-economic index. These groupings were not pre-specified. Robust standard errors clustered at the community level in parentheses. The D4 outcome is the average student z-score on the Kenya Certificate of Primary Education (KCPE) test. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table B6C—Round 1 treatment effects on key outcomes for spillover sample

	Control (1)	ITT (2)	TOT (3)	FDR *q*-val (4)
Panel A: Primary energy outcomes				
A1. Grid connected (%)	4.8	2.8	–	–
	[21.4]	(2.2)		
A2. Monthly electricity spending (USD)	0.09	0.02	0.10	–
	[0.60]	(0.05)	(0.24)	
Panel B: Additional energy outcomes				
B1. Electricity as main lighting source (%)	4.7	2.9	17.6*	0.364
	[21.1]	(2.2)	(10.1)	
B2. Number of appliance types owned	1.8	0.1	0.6	0.422
	[1.3]	(0.1)	(0.4)	
B3. Owns mobile phone (%)	85.0	1.7	12.6	0.449
	[35.7]	(2.3)	(10.9)	
B4. Owns radio (%)	53.6	-0.4	1.3	0.952
	[49.9]	(4.4)	(21.2)	
B5. Owns television (%)	18.2	3.0	21.5	0.449
	[38.6]	(3.9)	(17.9)	
B6. Owns iron (%)	3.9	2.4	14.7*	0.364
	[19.4]	(1.6)	(7.8)	
B7. Monthly kerosene spending (USD)	2.60	0.12	0.51	0.755
	[2.81]	(0.20)	(0.94)	
B8. Monthly total energy spending (USD)	7.99	0.25	0.95	0.925
	[11.20]	(0.88)	(4.22)	
B9. Solar home system as main lighting source (%)	14.3	1.1	9.0	0.631
	[35.0]	(2.4)	(11.2)	
Panel C: Primary economic outcomes				
C1. Household employed or own business (%)	38.3	-0.4	-1.7	0.977
	[39.3]	(2.7)	(13.6)	
C3. Total hours worked last week	45.8	0	-0.3	0.977
	[24.3]	(1.8)	(8.9)	
C4. Total asset value (USD)	894	5	111	0.977
	[1,007]	(127)	(610)	
C5. Per capita consumption of major items (USD)	138	-10	-45	0.947
	[139]	(13)	(61)	

(*Table continued on next page*)

(*Table continued from previous page*)

	Control	ITT	TOT	FDR q-val
	(1)	(2)	(3)	(4)
Panel D: Primary non-economic outcomes				
D1. Recent health symptoms index	0	0.03	0.24	0.950
	[1]	(0.09)	(0.44)	
D2. Normalized life satisfaction	0	-0.07	-0.29	0.947
	[1]	(0.08)	(0.40)	
D3. Avg. student test Z-score	0	0.08	0.54	0.947
	[1]	(0.12)	(0.62)	
D5. Political and social awareness index	0	0.02	0.24	0.947
	[1]	(0.07)	(0.32)	
Panel E: Mean treatment effects on grouped outcomes				
E1. Economic Index (C outcomes)	0	-0.03	-0.12	–
	[1]	(0.07)	(0.36)	
E2. Non-Economic Index (D outcomes)	0	0.06	0.50	–
	[1]	(0.09)	(0.43)	

Notes: Column 1 reports mean values in the control group, with standard deviations in brackets. Column 2 reports coefficients from separate ITT regressions in which the dependent variable (e.g., A1) is regressed on the high subsidy treatment indicator. The low and medium subsidy groups are excluded from these regressions. The sample size is 885 in these regressions, except for the D3 regression, in which the sample size is 619. Column 3 reports coefficients from separate TOT (IV) regressions in which household electrification status is instrumented with the three subsidy treatment indicators. The sample size is 1,328 for these regressions, except for the D3 regression, in which the sample size is 870. All specifications include pre-specified household, student, and community covariates. Column 4 reports the FDR-adjusted q-values associated with the coefficient estimates in column 3. FDR-adjusted q-values are computed for each outcome within the additional energy outcomes group (panel B), and for each outcome within the primary outcomes group (panels C and D combined). In panel E, we report mean treatment effects on outcomes grouped into an economic and non-economic index. These groupings were not pre-specified. Robust standard errors clustered at the community level in parentheses. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table B7—Benchmarking average monthly electricity consumption in kWh and USD

			Percentile			
		Mean	25th	50th	75th	N
Panel A: Newly connected households in sample						
R1 survey (2016)	kWh	7.9	0	3.6	22.0	506
	USD	2.29	0	1.98	3.30	
R2 survey (2017)	kWh	10.8	0	5.0	22.6	541
	USD	2.52	0	2.05	3.47	
Panel B: Baseline connected households in sample						
Baseline survey (2014)	kWh	62.0	12.7	42.9	65.4	149
	USD	10.57	3.41	6.82	11.39	
R1 survey (2016)	kWh	74.1	30.9	72.7	76.4	208
	USD	9.84	3.94	9.58	10.29	
R2 survey (2017)	kWh	72.3	14.8	59.9	79.6	195
	USD	9.97	2.84	7.58	11.37	
Panel C: Kenya Power customers (2014)						
Busia & Siaya	kWh	46.1	12.3	29.7	58.2	2,147
	USD	8.62	2.75	4.82	9.54	
Nationwide	kWh	85.1	18.6	40.5	87.6	111,084
	USD	16.62	3.39	6.03	15.18	
Kisumu	kWh	79.2	24.3	49	89.3	1,666
	USD	14.95	4.01	7.22	15.75	
Nairobi	kWh	189.9	30.3	72.8	178.6	15,577
	USD	39.33	4.71	12.07	34.8	

Notes: Panel A presents estimates of monthly electricity consumption in kWh and USD for newly connected households (i.e., treatment group households that were connected after the baseline survey). Panel B presents estimates for households that were already connected at baseline. Electricity consumption amounts are estimated using survey responses to the questions, "How much was the amount of your last monthly electricity bill?" for postpaid consumers, and "In the past three months, how much did you spend on top-ups" for prepaid consumers, and the 2014, 2016, and 2017 electricity rate structures. Panel C presents average monthly electricity consumption in kWh and USD for a random 10 percent sample of Kenya Power domestic accounts (i.e., mostly residential customers), based on electricity bills issued in 2014. In panels A and B, we use annual averages for certain components of the electricity bill (e.g., the Fuel Cost Charge, which fluctuates monthly). As a result, there are discrepancies between panels A and B in terms of conversions from kWh to USD. Kenya Shilling amounts are first adjusted to 2014 levels and then converted into U.S. dollars at the 2014 average exchange rate of 87.94.

Table B8A—Impact of randomized offers on hypothetical and actual take-up

	Stated WTP 1	Stated WTP 2	Actual take-up, experiment
	(1)	(2)	(3)
\$853 offer	-19.7***	-8.2***	
	(3.7)	(2.1)	
\$284 offer / T1: Low subsidy—29% discount	16.3***	6.0**	5.9***
	(3.4)	(2.5)	(1.5)
\$227 offer	14.3***	7.3***	
	(3.6)	(2.7)	
\$171 offer / T2: Medium subsidy—57% discount	24.1***	18.5***	22.9***
	(3.4)	(2.7)	(4.0)
\$114 offer	25.2***	19.7***	
	(3.5)	(2.9)	
Free offer / T3: High subsidy—100% discount	62.0***	87.5***	95.0***
	(2.9)	(2.2)	(1.3)
Age (years)	-0.4***	-0.2**	0.0
	(0.1)	(0.1)	(0.0)
Senior citizen=1	0.9	1.3	0.5
	(3.5)	(3.0)	(1.4)
Attended secondary school=1	15.6***	5.4**	3.8**
	(2.7)	(2.4)	(1.7)
Not a farmer=1	0.4	0.1	1.9
	(2.4)	(1.9)	(1.6)
Employed=1	2.3	1.2	1.1
	(2.2)	(1.9)	(1.3)
Has bank account=1	11.1***	11.0***	2.6
	(2.5)	(2.5)	(1.7)
Number of household members	1.3***	0.4	0.6***
	(0.4)	(0.3)	(0.2)
High-quality walls=1	9.1***	11.6***	3.5
	(2.7)	(2.3)	(2.1)
Number of chickens=1	0.7***	0.4***	0.1**
	(0.1)	(0.1)	(0.1)
Take-up in status quo (i.e., \$398) group	36.2	9.8	1.3
Mean of dependent variable	53.7	25.5	21.6
Observations	2,157	2,157	2,176
R^2	0.23	0.35	0.69

Notes: In column 1, the dependent variable is an indicator for whether the household accepted the hypothetical offer (i.e. randomly assigned price). In column 2, it is an indicator for whether the household accepted the hypothetical offer if required to complete the payment in six weeks. In column 3, it is an indicator for experimental take-up. All dependent variables are multplied by 100. Robust standard errors clustered at the community level in parentheses. All specfications include pre-specified community covariates including indicators for the county, market status, whether the transformer was funded and installed early on (between 2008 and 2010), community electrification rate at baseline, and community population. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table B8B—Impact of WTP offer on stated take-up of electricity connections

		Interacted variable		
	Baseline	High-quality walls	Has bank account	Attended secondary schooling
	(1)	(2)	(3)	(4)
\$853 offer	-8.2***	-6.0***	-8.0***	-5.4**
	(2.1)	(2.2)	(1.9)	(2.2)
\$284 offer / T1: Low subsidy—29% discount	6.0**	5.0**	4.9*	6.0**
	(2.5)	(2.4)	(2.6)	(2.4)
\$227 offer	7.3***	6.6**	7.2**	7.7***
	(2.7)	(2.8)	(2.8)	(2.7)
\$171 offer / T2: Medium subsidy—57% discount	18.5***	16.0***	16.6***	17.1***
	(2.7)	(2.7)	(2.9)	(2.7)
\$114 offer	19.7***	18.4***	15.0***	20.0***
	(2.9)	(3.2)	(2.9)	(2.9)
Free offer / T3: High subsidy—100% discount	87.5***	89.6***	89.6***	89.3***
	(2.2)	(2.3)	(2.1)	(2.2)
Interacted variable		7.9	5.6	7.2
		(5.3)	(4.8)	(5.9)
\$853 offer × interacted variable		-9.0	-4.1	-18.0***
		(6.4)	(7.6)	(6.1)
\$284 offer × interacted variable		6.4	5.5	0.0
		(8.5)	(7.3)	(8.8)
\$227 offer × interacted variable		4.6	0.6	-2.7
		(8.5)	(7.4)	(8.5)
\$171 offer × interacted variable		15.7*	9.9	11.4
		(8.3)	(7.8)	(9.9)
\$114 offer × interacted variable		8.5	25.1***	-2.1
		(8.5)	(8.4)	(9.2)
Free offer × interacted variable		-11.5*	-15.1**	-17.2**
		(5.9)	(5.9)	(6.6)
Take-up in status quo (i.e., \$398) group	9.8	9.8	9.8	9.8
Mean of dependent variable	25.5	25.5	25.5	25.5
Observations	2,157	2,157	2,157	2,157
R^2	0.35	0.36	0.36	0.35

Notes: The dependent variable is an indicator (multiplied by 100) for whether the household accepted the hypothetical offer if required to complete the payment in six weeks. Pre-specified household covariates include the age of the household head, indicators for whether the household respondent attended secondary school, is a senior citizen, is not primarily a farmer, is employed, and has a bank account, an indicator for whether the household has high-quality walls, and the number of chickens (a measure of assets) owned by the household. Pre-specified community covariates include indicators for the county, market status, whether the transformer was funded and installed early on (between 2008 and 2010), community electrification rate at baseline, and community population. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table B8C—Predictors of financial constraints in WTP questions

	(1)	(2)
\$853 offer	90.3***	91.4***
	(5.2)	(5.5)
\$398 offer / Existing fixed price	72.9***	75.9***
	(4.1)	(4.1)
\$284 offer / T1: Low subsidy—29% discount	70.3***	72.2***
	(3.3)	(3.4)
\$227 offer	65.9***	68.2***
	(3.7)	(3.8)
\$171 offer / T2: Medium subsidy—57% discount	52.7***	55.0***
	(3.3)	(3.4)
\$114 offer	52.9***	54.2***
	(3.3)	(3.4)
Age (years)		0.1
		(0.1)
Senior citizen=1		-3.5
		(5.2)
Attended secondary school=1		0.1
		(3.1)
Not a farmer=1		0.3
		(3.2)
Employed=1		0.4
		(2.9)
Has bank account=1		-10.7***
		(3.2)
Number of household members		-0.1
		(0.5)
High-quality walls=1		-12.5***
		(3.3)
Number of chickens=1		-0.2*
		(0.1)
Mean of dependent variable	52.4	52.5
Observations	1,184	1,159
R^2	0.25	0.27

Notes: In both columns, the dependent variable is an indicator (multiplied by 100) for whether the household first accepted the hypothetical offer (i.e. randomly assigned price) to connect to the grid, and then declined the hypothetical offer if required to complete the payment in six weeks. Robust standard errors clustered at the community level in parentheses. Asterisks indicate coefficient statistical significance level (2-tailed): * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$.

Table B9—Summary of randomly-assigned, hypothetical credit offers

Offer	Months	Upfront	Monthly	NPV at discount rate of 5%	15%	25%	*n*	Take-up: Time un-limited	6 week deadline
1	36	79.60	11.84	475.23	425.67	387.38	406	50.6%	38.3%
2	36	59.70	12.58	480.03	427.38	386.69	379	53.5%	38.9%
3	36	39.80	13.32	484.83	429.09	386.01	369	52.7%	39.6%
4	36	59.70	13.45	509.29	452.98	409.46	353	49.7%	39.1%
5	24	59.70	17.22	452.57	418.07	389.91	419	52.4%	40.2%
6	36	127.93	26.94	1028.26	915.48	828.34	363	52.7%	28.2%
Offer 1 to 5 (average)		59.70	13.68	480.39	430.64	391.89		52.0%	39.3%

Notes: During the baseline survey, each household was randomly assigned a hypothetical credit offer consisting of an upfront payment (ranging from $39.80 to $79.60), a monthly payment (ranging from $11.84 to $17.22), and a contract length (either 24 or 36 months). Respondents were first asked whether they would accept the offer, and then asked whether they would still accept if required to complete the upfront payment in six weeks. Figure 3, panel C plots the net present value and take-up results corresponding to offer 6 and the average for offers 1 to 5 (which are very similar), assuming a discount rate of 15 percent.

Table B10—Transformer problems documented in the study communities over a 14-month period (September 2014 to October 2015)

Row	Site ID	Group	Wave	Treated HHs	Connected	Metered	Blackout	Primary issue
1	1204	Treatment	2	15	Feb-15	May-15	4 months	Burnt out
2	1403	Treatment	1	15	Mar-15	Jul-15	1 month	Commissioning
3	1505	Treatment	2	1	Mar-15	May-15	1 month	Commissioning
4	2101	Treatment	1	0	n/a	n/a	8 months	Burnt out
5	2103	Treatment	1	0	n/a	n/a	4 months	Technical failure
6	2106	Treatment	1	15	Nov-14	Nov-14	8 months	Commissioning
7	2114	Treatment	1	8	Dec-14	Dec-14	12 months	Relocated by Kenya Power
8	2116	Treatment	1	14	Sep-14	May-15	2 months	Technical failure
9	2202	Treatment	1	1	Sep-14	Oct-14	1 month	Technical failure
10	2217	Treatment	1	13	Oct-14	Dec-14	1 month	Technical failure
11	2222	Treatment	1	3	Oct-14	Dec-14	4 months	Leaking oil
12	2303	Treatment	2	7	May-15	Jun-15	4 months	Technical failure
13	2406	Treatment	2	15	Apr-15	Jun-15	1 month	Burnt out
14	2503	Treatment	1	1	Oct-14	Oct-14	6 months	Burnt out
15	2506	Treatment	1	15	Dec-14	Feb-15	9 months	Commissioning
16	1103	Control	n/a	0	n/a	n/a	2 months	Technical failure
17	1109	Control	n/a	0	n/a	n/a	6 months	Burnt out
18	1203	Control	n/a	0	n/a	n/a	1 month	Technical failure
19	1205	Control	n/a	0	n/a	n/a	1 month	Technical failure
20	1405	Control	n/a	0	n/a	n/a	6 months	Burnt out
21	1410	Control	n/a	0	n/a	n/a	2 months	Relocated by Kenya Power
22	2103	Control	n/a	0	n/a	n/a	4 months	Burnt out
23	2115	Control	n/a	0	n/a	n/a	2 months	Technical failure
24	2212	Control	n/a	0	n/a	n/a	5 months	Burnt out
25	2220	Control	n/a	0	n/a	n/a	8 months	Burnt out
26	2304	Control	n/a	0	n/a	n/a	3 months	Stolen
27	2315	Control	n/a	0	n/a	n/a	3 months	Burnt out
28	2504	Control	n/a	0	n/a	n/a	4 months	Technical failure
29	2515	Control	n/a	0	n/a	n/a	4 months	Damaged by weather

Note: "Commissioning" refers to a situation in which the transformer (and related equipment) is installed but electricity is not being delivered.

Table B11—Costs of infrastructure construction associated with electricity connection projects

	Budgeted		Invoiced (*Panel A*) Observed (*Panel B*)			Difference	
	Total	Per HH	Total	Per HH	Allocation	Amount	%
Panel A: Project costs, budgeted and invoiced							
Local network	383,207	798	358,235	749	61.1%	-24,972	- 6.5%
Labor and transport	177,457	370	200,080	419	34.1%	+22,623	+12.7%
Service lines	15,812	33	27,684	58	4.7%	+11,873	+75.1%
Total cost	576,476	1,201	585,999	1,226	100.0%	+9,523	+1.7%
Panel B: Project materials, budgeted and observed							
Electricity poles	1,449	3.0	1,141	2.4	–	-308	-21.3%

Notes: In panel A, project costs are reported in USD and consist of administrative budgeted estimates and final invoiced amounts. "Local network" consists of high- and low-voltage electricity poles and cables. "Labor and transport" also includes design work and small contingency items. "Service lines" are typically single "drop-down" cables that connect households to an electricity line. Kenya Power metering costs and houshold wiring costs are not included in this summary. In total, the project involved roughly 101.6 km of new low-voltage lines. In panel B, we compare the budgeted number of electricity poles to the actual number of poles that were observed to have been installed.

Table B12—Detailed breakdown of labor and transport costs for nine projects (three contracts)

	Contract #1	Contract #2	Contract #3
Panel A: Labor costs (e.g., digging holes, installation, clearing bush, dropping service lines, etc.)			
Budgeted LV poles	40	107	62
Invoiced LV poles	38	98	76
Actual (counted) LV poles	39	92	60
Difference (Actual - Invoiced)	*+1*	*-6*	*-16*
Avg. labor cost per LV pole	*27.59*	*27.59*	*27.59*
Total LV poles labor	1,048	2,704	1,655
Budgeted stays	–	–	35
Invoiced stays	32	68	43
Avg. labor cost per stay	*19.22*	*19.22*	*19.22*
Total stays labor	615	1,308	827
Budgeted HV poles	–	–	6
Invoiced HV poles	12	5	6
Avg. labor cost per HV pole	*35.59*	*35.59*	*35.59*
Total HV poles labor	427	178	214
Additional labor	832	1,552	2,199
Total labor	2,922	5,742	4,895
Panel B: Transport costs (e.g., wood pole and other materials)			
Large lorries	2	4	4
Invoiced round-trip distance (km)	320	300	300
Google round-trip distance (km)	218	256	218
Difference (Actual - Invoiced)	*-102*	*-44*	*-82*
Avg. cost per km	*3.75*	*3.75*	*3.75*
Total large lorry transport	2,402	4,503	4,503
Small lorries	1	3	2
Invoiced round-trip distance (km)	250	250	250
Avg. cost per km	2.98	2.98	2.98
Total small lorry transport	745	2,234	1,490
Total transport	3,146	6,738	5,993
Budgeted labor and transport costs	6,126	12,708	8,956
Invoiced labor and transport costs	7,040	14,477	12,516
Difference (Invoiced - Budgeted)	*14.9%*	*13.9%*	*39.8%*
Projects	3	3	3
Households connected	18	38	22
Construction days	36	31	35

Notes: Based on the detailed invoice submitted to REA. "LV" denotes low-voltage and "HV" denotes high-voltage. Additional labor includes costs of bush clearing, tree cutting, signage, dropping service cables, and other expenses. Each large lorry is capable of transporting 30 poles. Each small lorry is capable of transporting 2.3 km of line materials.

Table B13—Predicting social surplus per household (SS) under different assumptions

		Experimental approach		Alternative approach		
	C	CS	SS	CS	SS	Key assumption(s)
Main estimates	739	147	-593	293	-446	
a) Income growth (*experimental approach*)	–	+139		–		Growth of 3 percent per annum over 30 years (based on figure 2, panel B).
Electricity consumption growth (*alternative approach*)	–	–		+365		Growth of 10 percent per annum over 30 years (see table 4, column 2, row 3).
b) No credit constraints for grid connections	–	+301		–		Stated WTP without time constraints (see figure 3, panel C)
c) No transformer breakdowns	–	+33		+37		Reduce transformer breakdowns from 5.4 to 0 percent (see appendix table B10).
d) No connection delays	–	+46		+52		Reduce waiting period from 188 to 0 days (see appendix figure A1).
e) No construction cost leakage	-157	–		–		Decrease total construction costs by 21.3 percent (see appendix table B11).
f) Including baseline connected households	**–**	**+92**		**+87**		**Net effect of incorporating a weighted average consumer surplus (based on table 4, column 3, row 3).**
Ideal scenario	582	757	175	834	253	

Notes: Main estimates of C, CS, and SS correspond to the values shown in figure 3, panel B (for the experimental approach), and table 4, column 1, row 3 (for the alternative approach). Row f incorporates consumer surplus from baseline connected households (roughly 5.5 percent of community households). Specifically, these values reflect the net impact on the bottom row of incorporating a weighted average consumer surplus, using the estimate in table 4, column 3, row 3 as a proxy for the consumer surplus from baseline connected households.

Appendix C

This appendix contains the three pre-analysis plans referenced in the main text. The pre-analysis plans are also available at http://www.socialscienceregistry.org/trials/350.

Pre-analysis plan A
"The demand for and costs of supplying grid connections in Kenya"

AEA RCT Title: "Evaluation of Mass Electricity Connections in Kenya"

RCT ID: AEARCTR-0000350

Principal Investigators: Eric Brewer, Kenneth Lee, Edward Miguel, and Catherine Wolfram

Date: 30 July 2014

Summary: This document outlines the plan for analyzing the demand for and costs of supplying household electricity connections in rural Kenya. The proposed analysis will take advantage of a field experiment in which randomly selected clusters of rural households were offered an opportunity to connect to the national grid at subsidized prices. This pre-analysis plan outlines the regression specifications, outcome variables, and covariates that will be considered as part of this analysis. We anticipate that we will carry out additional analyses beyond those included in this plan. This document is therefore not meant to be comprehensive. The overall research project will also include an impact evaluation of electricity connections that will be carried out in 2015 or 2016, upon completion of the endline survey round. For this portion of the project, we will register an additional pre-analysis plan at a later date, in either 2015 or 2016.

I. Introduction

Electrification has long been a benchmark of development, yet over two-thirds of the population of Sub-Saharan Africa lives without access to electricity. In June 2013, President Obama announced the Power Africa initiative, making energy access a top priority among six partner countries in Africa, including Kenya. In light of this initiative, and others being implemented by the World Bank and the UN General Assembly, there is considerable need for rigorous research to inform the effective scale-up of energy access programs in developing countries.

In this project, we have identified a unique opportunity to increase access to on-grid energy in Kenya. Since 2007, Kenya's Rural Electrification Authority (REA) has rapidly expanded the national grid, installing electricity distribution lines and transformers across many of the country's rural areas. Connectivity, however, remains low. While roughly three-quarters of the population is believed to live within 1.2 kilometers of a low voltage line, the official electrification rate is under 30%. In related work, we find that in regions that are technically covered by the grid, half of the unconnected households are no more than 200 meters from a low-voltage line.

We believe that the primary barrier to connecting these "under grid" households is the prohibitively high connection fee faced by rural households. The current connection price of KSh 35,000 ($412) may not be affordable for poor, rural households in a country where the GNI per capita (PPP) is $1,730. Despite this fact, Kenya's monopoly distribution company, Kenya Power, has recently proposed increasing the price to KSh 75,000 due to cost considerations.[1]

In general, little is known about the demand for electricity in rural areas, both initially and over time. Specifically, how many more households would opt to connect if the fee were,

[1] In March 2014, Kenya Power, the national utility, stated that it will continue to charge eligible customers KSh 35,000 for single-phase power connections, as long as the cost of connection does not exceed KSh 135,000 ($1,588), inclusive of VAT.

for example, KSh 25,000 ($294), KSh 15,000 ($176), or even KSh 0? How much power would households consume if they did connect, now and in the future? And once households are connected, do the social and economic benefits of access to modern energy in rural areas outweigh the costs?

In the coming years, REA will explore the feasibility of initiating a long-term, last-mile household connection program involving discounted connection fees for households and small businesses located close to existing REA electricity transformers. In order to evaluate this potential program, we have partnered with REA to conduct a randomized evaluation of grid connections involving roughly 2,500 households in rural Western Kenya.

The principal objectives of this study are twofold:

1. To trace out the demand curve for electricity connections, and in addition, to estimate the economies of scale in costs associated with spatially grouping connections together.

2. To measure the social and economic impacts of electrification, including schooling outcomes for children, energy use, income and employment, among other outcomes.

This pre-analysis plan outlines our strategy to address the first objective. The analysis on the impacts of the intervention will be carried out in 2015 and 2016, upon completion of the midline and endline survey rounds. The pre-analysis plan for the second stage of this project will therefore be registered at a later date, in either 2015 or 2016.

The remainder of this document is organized as follows. Section II provides a brief background on the existing literature on the demand for electricity connections. Section III provides a brief overview of the experimental design. Finally, Sections IV and V outline the main estimating equations that will be used in our analysis of both the demand for and costs of supplying electricity connections.

II. Brief literature review

In recent years, there has been a growing literature examining the demand for electricity connections in developing countries. The methods utilized in these studies range from contingent valuation approaches (see, e.g., Abdullah and Jeanty 2011) to randomized encouragement designs, where households are offered vouchers or subsidies to connect to the electricity network at a discounted price. Bernard and Torero (2013), for example, distribute two levels of randomized vouchers (10% and 20% discounts) to encourage household grid connections in Ethiopia, where the connection price ranges from $50 to $100, depending on the household's distance to the nearest electrical pole. Similarly, Barron and Torero (2014) utilize two levels of randomized vouchers (20% and 50% discounts) in El Salvador, where the connection price (in the study setting) is $100.

There is also an engineering literature simulating the costs of extending the grid to rural areas in developing countries. Parshall et al. (2009), for example, apply a spatial electricity planning model to Kenya and find that "under most geographic conditions, extension of the national grid is less costly than off-grid options." Zvoleff et al. (2009) examine the costs associated with extending the grid across various types of settlement patterns, demonstrating the potential for non-linearities in costs.

While our study is closely related to the earlier randomized encouragement designs, our objective is to evaluate the demand for electricity connections at randomized prices, as well as provide experimental evidence on the cost economies of scale associated with grouping connections together spatially.

III. Overview of project

1. Experimental design

Our experiment takes place across 150 "transformer communities" in Western Kenya. Each transformer community is defined as the group of all households located within 600 meters of a central electricity distribution transformer. In Kenya, all households within

600 meters of a transformer are eligible to apply for an electricity connection. In each transformer community, we have enrolled roughly 15 randomly selected unconnected households. In total, our study will involve roughly 2,250 unconnected households.

On 23 April 2014, our sample of transformer communities was randomly divided into treatment and control groups of equal size (75 treatment, 75 control). Each of the 75 treatment communities were then randomly assigned to one of three treatment arms (i.e. subsidy groups). These subsidies were designed to allow households to connect to the national power grid at relatively low prices (compared to the current connection price of KSh 35,000 or $412). In addition, each household accepting an offer to be connected as part of the study would receive a basic household wiring solution ("ready-board") at no additional cost. Each ready-board provides a single light bulb socket, two power outlets, and two miniature circuit breakers (MCBs).

The treatment and control groups are characterized as follows:

<u>A. High-value treatment arm</u>

25 communities. KSh 35,000 ($412) subsidy and KSh 0 ($0) effective price. This represents a 100% discount on the current price.

<u>B. Medium-value treatment arm:</u>

25 communities. KSh 20,000 ($235) subsidy and KSh 15,000 ($176) effective price. This represents a 57% discount on the current price.

<u>C. Low-value treatment arm:</u>

25 communities. KSh 10,000 ($118) subsidy and KSh 25,000 ($294) effective price. This represents a 29% discount on the current price.

<u>D. Control group:</u>

75 communities. No subsidy and KSh 35,000 ($412) effective price. There is no discount offered to households in the control group.

Within each treatment community, all enrolled and unconnected households would receive the same subsidy offer. After receiving the subsidy offer, treatment households would be given eight weeks to accept the offer and deliver the required payment to REA. At the end of this eight-week period, field enumerators would visit each household to verify that the required payment has been made to REA. Electricity connections are delivered once these verifications are complete. The collection of take-up responses comprises the main data set for the analyses outlined in this pre-analysis plan.

Once payments are verified, REA would hire its own contractors to deliver the connections within a period of four to six weeks. In order to economize on its own delivery costs, REA would connect all of the required connections in each community at the same time. REA would also group anywhere from two to four neighboring communities together, in order to further economize on transportation costs.

The first set of randomized offers were delivered in early-May and expired in early-July. The second set of randomized offers will be delivered in late-July and will expire in late-September. Our field enumerators began collecting take-up data on 4 July 2014. The full round of data collection will continue through the end of October 2014. As a result, it is expected that the final version of the data set for this analysis will be available in November 2014.

Data collection began before this document was uploaded to the AEA RCT registry website. In anticipation of this delay, we posted a document to our registered trial on 2 July 2014 titled "A note on pre-analysis plans" in order to describe how the investigators would be prohibited from accessing any data until a pre-analysis plan had been uploaded to the registry website.

2. Power calculations

At the beginning of this project, we knew little about the demand for electricity connections at various prices. We therefore made a set of assumptions on how take-up would vary at four different levels of prices. Taking into account our budgetary constraints, we designed the study to detect differences in take-up at these pricing levels, based on our set of ex-ante assumptions. In addition, we took into consideration the level of take-up that we would need in our future analysis on the social and economic impacts of electrification. These assumptions are outlined in Table 1.

Table 1: Ex-ante take-up assumptions

	Communities	Households (n)	Assumed take-up range
A. High-value arm ("High")	25	375	90 - 95%
B. Medium-value arm ("Medium")	25	375	40 - 50%
C. Low-value arm ("Low")	25	375	15 - 25%
D. Control group ("Control")	75	1,125	0 - 5%
Total	150	2,250	

Table 2: Communities required in each arm to detect differences with 80% power

Comparison	Description	Required size of each arm	Actual size of each arm
A vs. B	High vs. Med.	3 - 5	25
A vs. C	High vs. Low	2	25
A vs. D	High vs. Control	1 - 2	25 (High), 75 (Control)
B vs. C	Med. vs. Low	6 - 27	25
B vs. D	Med. vs. Control	3 - 5	25 (Med), 75 (Control)
C vs. D	Low vs. Control	6 - 26	25 (Low), 75 (Control)

In Table 2, we report the total number of communities required to detect differences ($\alpha = 0.05$) between groups with 80% power. For example, in the comparison of groups B (medium-value treatment arm) and C (low-value treatment arm), we expect that we will need 6 to 27 communities in each treatment arm (the actual size of each arm is 25 communities).[2] We assume an intracluster correlation coefficient of 0.1 within communities. In our design, we included a large number of high-value treatment communities in order to increase our statistical power to estimate the social and economic impacts of electrification (our second objective). Based on these assumptions, we expect that we are

[2]Since we had assumed a range of values for our assumptions on take-up, we report a range of values for the required size of each arm. For example, if take-up is 50% and 15% for groups B and C, respectively, we would require only 6 communities in each arm. However, if take-up is 40% and 25% for groups B and C, respectively, we would require 27 communities.

sufficiently powered, based on our ex-ante assumptions on take-up.

3. Data

This analysis will utilize four data sets: (1) Data on household take-up decisions; (2) Data on actual costs of supplying household connections; (3) Data on community-level characteristics; and (4) Household-level baseline survey data from the Living Standards Kenya (LSK) survey. The survey instrument is included in the Appendix.

IV. Analysis plan - Demand

The primary objective of this analysis is to estimate the demand for electricity connections, or in other words, the willingness of individual households to pay for a quoted price of an electricity connection. We will follow the procedure: (1) Estimate a non-parametric regression of household take-up on various subsidy levels. (2) Test for linearity: If we cannot reject linearity, we will estimate a linear regression of take-up on the effective connection price. If we can reject linearity, we will focus on the non-parametric estimation for the remainder of the analysis. (3) Estimate heterogeneous effects. (4) Plot the demand curve and compare these results to our contingent valuation results.

1. Non-parametric regression

We will begin by estimating the main equation:

$$y_{ic} = \alpha_0 + \alpha_1 T_c^{low} + \alpha_2 T_c^{mid} + \alpha_3 T_c^{high} + X_c' \gamma + \epsilon_{ic} \quad (1)$$

where y_{ic} is a binary variable reflecting the take-up decision for household i in transformer community c.[3] The binary variables T_c^{low}, T_c^{mid}, and T_c^{high} indicate whether community c was randomly assigned into the low-value, medium-value, or high-value treatment arms, respectively. Following Bruhn and McKenzie (2009), we include a vector of community-level characteristics, X_c, containing the variables used for stratification dur-

[3] Refer to Section IV Part 3 for further details on the dependent variable.

ing randomization.[4] Standard errors will be clustered at the community level.

Equation (1) will be the primary equation that we estimate in our demand-side analysis. As a robustness check, we will also estimate the equation:

$$y_{ic} = \alpha_0 + \alpha_1 T_c^{low} + \alpha_2 T_c^{mid} + \alpha_3 T_c^{high} + X_c' \gamma + X_{ic}' \lambda + \epsilon_{ic} \tag{2}$$

where X_{ic} is a vector of household-level characteristics.[5] X_{ic} will include standard control variables that not only have predictive effects but may also serve as sources of heterogeneity in take-up.

We will also assess whether treatment and control households are balanced at baseline in terms of household characteristics. In addition to X_{ic}, we may also choose to control for any covariates that are both unbalanced at baseline and relevant for electricity take-up.

In equations (1) and (2), the baseline (i.e. $T_c^{low} = T_c^{mid} = T_c^{high} = 0$) estimates household take-up under the status-quo pricing policy (i.e. take-up when the price of an electricity connection faced by the rural household is KSh 35,000). α_1, α_2, and α_3 capture the incremental effects (over the baseline) on take-up of the low-value, medium-value and high-value subsidies, respectively. Since the randomized subsidies will lower the effective price of an electricity connection, we expect that our experiment will result in positive and statistically significant α-coefficients.

2. Testing for linearity

We are interested in testing for linearity in equation (1). We will use an F-test to assess the null hypothesis:

$$H_0\colon \frac{(\alpha_3 - \alpha_2)}{15} = \frac{(\alpha_2 - \alpha_1)}{10} = \frac{(\alpha_1 - \alpha_0)}{10}$$

[4] Refer to Section IV Part 4 for further details on the components of X_c.

[5] Refer to Section IV Part 4 for further details on the components of X_{ic}.

against the alternative hypothesis that the slope in between the various take-up points is unequal. If we cannot reject linearity in an F-test, we will also estimate the equation:

$$y_{ic} = \beta_0 + \beta_1 p_c + X_c' \gamma + \epsilon_{ic} \tag{3}$$

where p_c is the effective price of an electricity connection faced by households in community c.[6]. Standard errors will again be clustered at the community level. As in equation (2), we will similarly check robustness by including the vector X_{ic}.

If we can reject linearity in an F-test, it will be of interest to understand how take-up changes when moving across different subsidy levels. In a similar experiment conducted in El Salvador, Barron and Torero (2014) find that the effects of a relatively low subsidy (20%) and a relatively high subsidy (50%) are similar. This is taken to suggest that either the demand for connections is inelastic (in the price range offered), or that the subsidies affect take-up through alternative channels.[7] Given this unusual result, we will focus on equation (1) and test the hypothesis that:

$$H_0\text{: } \alpha_1 = \alpha_2$$

against the alternative that the higher-value subsidy has a larger effect on take-up compared to the lower-value subsidy (i.e. H_1: $\alpha_2 > \alpha_1$). We will conduct a similar test for each of the pairwise combinations listed in Table 2.

3. Two measures of take-up

We may find that some of the treatment households decided that they would like to accept the offer, but are unable to complete the full payment within the eight-week period. We may therefore have two measures of take-up:

[6] For example, in a high-subsidy treatment community, the subsidy amount is equal to the current price of an electricity connection and the effective price faced by households is 0 KSh (i.e. $p_c = 0$)

[7] For example, Barron and Torero propose that a subsidy may raise awareness that electrification is possible, resulting in higher take-up.

1. Actual take-up (y_{ic}^1): Binary variable indicating whether treatment household ic accepted the offer and completed the required payment within eight weeks.

2. Intended take-up (y_{ic}^2): Binary variable indicating whether treatment household ic intended to accept the offer, and began to make payments, but was unable to complete the full payment within eight weeks.

Our primary outcome of interest, however, will be the actual take-up captured by y_{ic}^1.

4. Covariate vectors X_c and X_{ic}

There are two sets of covariates in equations (1), (2), and (3). X_c is a vector of community-level characteristics and X_{ic}, which will mainly be used in robustness checks, is a vector of household-level characteristics. X_c will primarily include the stratification variables that were used during randomization.[8] The list of X_c variables will include:

1. County indicator: Binary variable indicating whether community c is in Busia or Siaya. This was used as a stratification variable during randomization.

2. Market status: Binary variable indicating whether the total number of businesses in community c is strictly greater than the community-level mean across the entire sample. We use this definition to define which communities could be classified as "markets" relative to the others. This was used as a stratification variable during randomization.

3. Transformer funding year: Binary variable indicating whether the electricity transformer in community c was funded "early" (i.e. in either 2008-09 or 2009-10). This was used as a stratification variable during randomization.

4. Electrification rate: Residential electrification rate in community c.

5. Community population: Estimated number of people living in community c.

X_{ic} will include a set of household-level variables that not only have predictive effects but may also serve as sources of heterogeneity in take-up. The survey from which we will obtain this data is attached in the Appendix. For example, it is possible that take-up will vary depending on household size, household wealth, or the education level and employment type of the survey respondent. In the majority of cases, the survey respondent

[8] The collection of this data is described in further detail in Lee et al. (2014).

is either the household head or the spouse of the household head. The list of X_{ic} variables will include (*LSK question numbers in parentheses*):

1. Household size ($a1$): Number of people living in household ic.

2. Household wealth indicator - Walling material ($c1c$): Binary variable indicating whether the walls of household ic can be considered "high quality" (i.e. made of brick, cement, or stone).

3. Household wealth indicator - Chickens ($d9a$): Number of chickens owned by household ic.

4. Age of respondent in years ($a4c$)

5. Education of respondent ($a5b$): Binary variable indicating whether respondent ic has completed some level of secondary education.

6. Farming as primary occupation of respondent ($a5c$): Binary variable indicating whether the primary occupation of respondent ic is farming.

7. Access to financial services of respondent ($g1a$): Binary variable indicating whether respondent ic uses a bank account.

8. Business or self employment activity of respondent ($e1$): Binary variable indicating whether the respondent (or the respondent's spouse) in household ic engages in any business or self-employment activities.

9. Senior household ($a4c$): Binary variable indicating whether respondent ic is over 65 years old.

5. Heterogeneous effects

We are interested in understanding how take-up varies across several important socio-economic dimensions. For example, will take-up depend on community characteristics? Will it be higher for households that are located in more electrified communities or in market centers? Alternatively, will take-up depend on individual characteristics? Will it be higher for the more educated households, or those that are engaged in more "entrepreneurial activities"? In order to answer these questions, we will estimate heterogeneous effects along a number of dimensions, captured in the vectors X_c and W_{ic} (which is a subset of X_{ic}):

1. County indicator (X_c)
2. Market status (X_c)
3. Transformer funding year (X_c)
4. Electrification rate (X_c)
5. Community population (X_c)
6. Household wealth indicator - Walls (W_{ic})
7. Education of respondent (W_{ic})
8. Farming as primary occupation of respondent (W_{ic})
9. Access to financial services of respondent (W_{ic})
10. Business or self employment activity of respondent (W_{ic})
11. Senior household (W_{ic})

We will estimate heterogeneous effects by adding interactions between the treatment variables and the vectors X_c and W_{ic} to equations (1), (2), and (3). We will also carry out additional analyses, depending on the types of heterogeneous effects that we estimate. For example, if we find that take-up is higher in communities with higher electrification rates, we may explore whether there are any "bandwagon" effects, as in Bernard and Torero (2013), by focusing on the interaction between the treatment and community electrification variables. Since we do not know the nature of these heterogeneous treatment effects, it is not possible to fully specify all of the potential analyses in this document.

6. Comparison of contingent valuation to revealed preference results

During the LSK survey round, conducted between February and July 2014, we asked respondents from unconnected households whether they would be hypothetically willing to connect to the national grid at a randomly selected price (see questions $f16b$ and $f16c$ in Appendix). These amounts were randomly drawn from the following set of prices:

$$\text{Hypothetical Price} \in \{0, 10000, 15000, 20000, 25000, 35000, 75000\}$$

This question was followed by an additional hypothetical question asking the respondent whether they would accept an offer at this price if they were given six weeks to complete the payment.[9]

In comparison, there were four effective prices (randomized at the community-level) in our experimental design:

$$\text{Effective Price} \in \{0, 15000, 25000, 35000\}$$

By making comparisons between these two measures of take-up at similar levels of prices, we will test whether we could reject equal demand (in terms of contingent valuation and revealed preferences). In addition, we will plot various demand curves, with take-up plotted along one axis and the effective (or hypothetical) price plotted along the other. Finally, we will run contingent valuation regressions using the same specifications and covariates as those described in Section IV, Parts 1, 2, and 6.

V. Analysis plan - Costs

The secondary objective of this analysis is to characterize how connection costs decrease with the number of neighboring households that choose to connect at the same time.[10]

1. Potential for economies of scale in costs

Given that rural households are often located in remote areas, the cost of supplying an electricity connection to an individual household can be very high. This is due to the high cost of transportation and the necessity of building additional low-voltage lines. However, significant economies of scale could be achieved by connecting multiple households

[9] In our experimental design, treatment households were given eight weeks to complete the payment. This change was made at the request of REA, after we had already launched our baseline survey round. In this hypothetical question, we do not believe that providing an additional two weeks would have influenced the responses.

[10] We make a distinction between the *price* of an electricity connection, which is the fixed price of an electricity connection faced by households, and the *cost* of an electricity connection, which is the physical cost of supplying the electricity connection faced by the utilities.

at the same time. In a related paper, we use the current costs of materials to estimate that the incremental cost of supplying an electricity connection to a single household 200 and 100 meters away from a low-voltage line is \$1,940 and \$1,058, respectively, inclusive of material and transportation costs, as well as a 25% contractor markup (Lee et al. 2014).

While this cost is extremely high, it is desirable from the perspective of the supplier to connect spatially-clustered groups of households at the same time. For example, when two neighboring households are connected along the same length of line, the above per household costs are projected to fall by roughly 47%, to \$1,021 and \$580, respectively.

2. IV approach to estimating economies of scale in costs

In our experimental design, randomized subsidies are assigned at the community level. In addition, there are three levels of subsidies. We expect that different levels of subsidies—low, medium, and high—will create variation in the number of households that choose to apply for electricity at the same time. For example, larger numbers of applicants should be observed in the high-subsidy communities (where households pay 0 KSh), and smaller numbers of applicants should be observed in the low-subsidy communities (where households pay 25,000 KSh).

We can therefore estimate the community-level construction cost, Γ_c, as a function of the number of connected households in the community, M_c, using the randomized community-level subsidy amounts, Z_c^{low}, Z_c^{mid}, and Z_c^{high}, as instruments for M_c.[11] In order to allow for the possibility of non-linearities in costs, we will include higher-order polynomials in our estimation of Γ_c. Specifically, we will estimate an instrumental variables regression using the equations:

$$M_c = \delta_0 + \delta_1 Z_c^{low} + \delta_2 Z_c^{mid} + \delta_3 Z_c^{high} + V_c' \mu + \nu_c \tag{4}$$

[11] Refer to Section V Part 3 for additional information on how we plan to construct the variable Γ_c.

$$M_c^2 = \delta_0 + \delta_1 Z_c^{low} + \delta_2 Z_c^{mid} + \delta_3 Z_c^{high} + V_c' \mu + \nu_c \tag{5}$$

$$M_c^3 = \delta_0 + \delta_1 Z_c^{low} + \delta_2 Z_c^{mid} + \delta_3 Z_c^{high} + V_c' \mu + \nu_c \tag{6}$$

$$\Gamma_c = \pi_0 + \pi_1 M_c + \pi_2 M_c^2 + \pi_3 M_c^3 + V_c' \mu + \eta_c \tag{7}$$

where the first-stage equations (4), (5), and (6) estimate the effects of the treatment variables on the number of applicants, and the second-stage equation (7) estimates the effect of higher-order polynomials of the number of connected households on the community-level cost. Since there are multiple endogeneous variables in this framework, equations (4), (5), and (6) will be estimated jointly. V_c is a vector of community-level characteristics that will be relevant in this regression.[12] ν_c and η_c are error terms.

We will take the derivative of our estimates in equation (7) in order to uncover different points along the marginal cost curve. We will plot these points to sketch out a marginal cost curve, with the number of connected households on the horizontal axis and the marginal cost on the vertical axis. We will also expand equations (4) through (7) by interacting the Z_c and M_c variables with the V_c vector to explore any potential heterogeneous effects.

We should note that this analysis is highly speculative. We have not carried out any power calculations because we do not have baseline data on the community-level costs of household electrification. Furthermore, our ability to identify the desired effects will depend on the specified functional forms. If we estimate linear relationships in both stages, we will focus only on estimating equation (4) in the first-stage and substitute equation (7) with the equation:

$$\Gamma_c = \pi_0 + \pi_1 M_c + V_c' \mu + \eta_c \tag{8}$$

[12] Refer to Section V Part 4 for further details on the components of V_c.

In addition, we may pursue additional analyses, depending on the nature of the cost data that we eventually receive.

<u>3. Constructing the variable Γ_c</u>

Through our partnership with REA, we will collect actual cost invoices related to the connections that are delivered as a part of this study. Specifically, we will be provided with an itemized list of costs (e.g. cost of low-voltage lines, cost of service lines, cost of transportation etc.), as well as the design drawings detailing the planned locations of electricity poles. Using these data, we will work with REA to determine the total construction cost for each community.

<u>4. Covariate vector V_c</u>

V_c will include variables that should have an impact on construction costs, including all of the community-level variables in X_c, in addition to a community distance and land gradient variables. The list of V_c variables will include:

1. County indicator
2. Market status: This may approximate community density or the pre-existing coverage of the local low-voltage network.
3. Transformer funding year
4. Electrification rate: This should approximate the pre-existing coverage of the local low-voltage network. Higher electrification rates (and more local low-voltage network coverage) should decrease construction costs.
5. Community population
6. Distance from REA warehouse: Travel distance (in kilometers) between community c and the primary REA warehouse located in Kisumu where the construction materials are stored. Longer travel distances should increase construction costs.
7. Terrain or land gradient: We will use two different measures of terrain or land gradient. Dinkelman (2011) identifies land gradient as a major factor contributing to the costs of electrification. In flatter areas, the soil tends to be softer, making it cheaper to lay power lines and erect transmission poles. Our primary community-level land gradient variable will therefore be constructed using the same methodology as Dinkelman (2011). Specifically, we will use the 90-meter Shuttle Radar Topog-

raphy Mission (SRTM) Global Digital Elevation Model (available at www.landcover.org) to access elevation data and then construct measures of the average land gradient for each transformer community.[13] Our secondary community-level land gradient variable will be the variance in the distribution of altitudes collected across the entire population of geo-tagged buildings for each transformer community.[14]

References

Abdullah, Sabah and P. Wilner Jeanty. 2011. Willingness to pay for renewable energy: Evidence from a contingent valuation survey in Kenya. *Renewable and Sustainable Energy Reviews* 15: 2974-2983.

Barron, Manuel and Maximo Torero. 2014. Short Term Effects of Household Electrification: Experimental Evidence from Northern El Salvador.

Bernard, Tanguy and Maximo Torero. 2013. Bandwagon Effects in Poor Communities: Experimental Evidence from a Rural Electrification Program in Ethiopia.

Bruhn, Miriam and David McKenzie. 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics* 1(4): 200-232.

Dinkelman, Taryn. 2011. The Effects of Rural Electrification on Employment: New Evidence from South Africa. *American Economic Review* 101(December 2011): 3078-3108.

Lee, Kenneth, Eric Brewer, Carson Christiano, Francis Meyo, Matthew Podolsky, Javier Rosa, Catherine Wolfram, and Edward Miguel. 2014. Barriers to Electrification for "Under Grid" Households in Rural Kenya. NBER Working Paper 20327. National Bureau of Economic Research, Cambridge, MA. http://www.nber.org/papers/w20327.

Parshall, Lily, Dana Pillai, Shashank Mohan, Aly Sanoh, and Vijay Modi. 2009. National electricity planning in settings with low pre-existing grid coverage: Development of a spatial model and case study of Kenya. *Energy Policy* 37: 2395-2410.

Zvoleff, Alex, Ayse Selin Kocaman, Woonghee Tim Huh, and Vijay Modi. 2009. The impact of geography on energy infrastructure costs. *Energy Policy* 37(10): 4066-4078.

[13]Each transformer community is defined as all of the buildings within a 600 meter radius of a central electricity distribution transformer, as defined in Lee et al. (2014).

[14]Usage of this secondary definition of land gradient will depend on whether we can verify that our altitude records (taken using the GPS application on Android tablets) are relatively accurate.

Pre-analysis plan B

"The Economic and Social Impacts of Electrification: Evidence from Kenya"[1]

AEA RCT Title: "Evaluation of Mass Electricity Connections in Kenya"

RCT ID: AEARCTR-0000350

Principal Investigators: Kenneth Lee, Edward Miguel, and Catherine Wolfram (University of California, Berkeley)

Date: 15 September 2016

Summary: This document outlines the plan for analyzing a dataset consisting of information on the living standards of roughly 4,000 households in Western Kenya, including nearly 500 households that previously benefited from a randomized household electrification program. The goal of this study is to estimate the economic and social impacts of household electricity connections. This document lays out the main regression specifications and outcome variable definitions that we intend to follow. However, we anticipate that we will carry out additional analyses beyond those included in this document. This document is therefore not meant to be comprehensive or to preclude additional analyses.

[1] We are grateful to Susanna Berkouwer for assistance in preparing this document. This research is supported by the Berkeley Energy and Climate Institute, the Blum Center for Developing Economies, the Center for Effective Global Action, the Development Impact Lab (USAID Cooperative Agreements AID-OAA-A-13-00002 and AIDOAA-A-12-00011, part of the USAID Higher Education Solutions Network), the International Growth Centre, the U.C. Center for Energy and Environmental Economics, the Weiss Family Program Fund for Research in Development Economics, the World Bank, and a private donor. Corresponding author: Edward Miguel (emiguel@berkeley.edu).

1. Introduction

1.1 Summary

Universal access to modern energy has become a top priority for policymakers, nongovernmental organizations, and international donors across Sub-Saharan Africa. In Kenya, nearly $600 million has been invested in extending the grid to rural areas since 2008. While there is now widespread grid coverage, the national household electrification rate remains relatively low. Kenya is currently pursuing a strategy of last-mile connections for “under grid” households in order to reach universal access to electricity by 2020. Given the high cost of subsidizing mass connections, however, there is a need for better understanding of the impacts of rural electrification. In this study, we will provide experimental evidence on the impacts of household electrification across a range of economic and social outcomes in Western Kenya. We will also examine the impacts of grid connections on neighboring households to better understand possible spillovers.

Between 2013 and 2015, we implemented a field experiment in which electricity connection vouchers (worth varying amounts) were randomly assigned to clusters of rural households in Western Kenya. Households accepting these vouchers were then connected to the national grid, in cooperation with Kenya’s Rural Electrification Authority (REA) and Kenya Power, the main electricity distribution company. As a result of this experiment, it is possible to perform a randomized evaluation of household grid connections. The study focuses on household survey data from baseline and follow-up surveys of 2,294 “main sample” households, as well as survey data from a follow-up survey of roughly 1,200 “secondary sample” households.[2]

1.2 Experimental design and steps

In this section, we describe the experimental design. For further details, see Lee et al. (2016) at http://dx.doi.org/10.1016/j.deveng.2015.12.001, Lee, Miguel, and Wolfram (2016a) at http://dx.doi.org/10.1257/aer.p20161097, and Lee, Miguel, and Wolfram (2016b) at http://www.nber.org/papers/w22292.

Step 1: In July 2013, we collaborated with REA to identify a list of 150 rural “transformer communities” that would form a representative sample of communities recently connected to

[2] The distinction between “main” and “secondary” sample households is described in Section 1.3.

the electrical grid in Busia and Siaya, two counties in Western Kenya. Each community is defined as all of the structures that were located within 600 meters of a central transformer.

Step 2: Between September 2013 and December 2013, we visited each community and geo-tagged over 13,000 structures, capturing the universe of unelectrified households that could potentially be connected to the national grid.

Step 3: Using these data as a sampling frame, we randomly sampled 2,504 households, consisting of 2,294 households that were unconnected at baseline and 205 households that were connected to the grid at baseline. The regressions described in Section 2.2 will focus on the group of 2,294 households. We use data from the sample of 210 connected households mainly for descriptive purposes, for example, to compare characteristics of households that had already connected without our subsidy to households that later connected with a subsidy. Between February and August 2014, we administered a detailed survey of each household, capturing baseline measures of living standards ("Living Standards Kenya (LSK) Survey – Baseline (2014)").

Step 4: In April 2014, we randomly assigned the 150 communities into four groups: (1) "High-subsidy" (or 100% discount) arm with 25 communities, resulting in an effective price of $0; (2) "Medium-subsidy" (or 57% discount) arm with 25 communities, resulting in an effective price of $171; (3) "Low-subsidy" (or 29% discount) arm with 25 communities, resulting in an effective price of $284, and (4) "No subsidy" or control group (effective price of $398 plus wiring) with 75 communities.

Step 5: After distributing the electricity connection subsidies, we facilitated the construction of grid infrastructure to connect the 478 unconnected households that accepted the randomized offer. The first household was metered in September 2014, the average connection time was seven months, and the final household was metered over a year later, in October 2015.

Step 6: In May 2016, we launched a follow-up survey round targeting all 2,504 households enrolled during the baseline round, in addition to roughly 1,500 newly enrolled households from the same transformer communities. This new sample of 1,500 households will consist of roughly 1,200 households unconnected at baseline (i.e., those that were observed to be unconnected at the time of the baseline census), and roughly 300 connected households. The secondary sample regressions described in Section 2.4 will focus on the group of 1,200 households unconnected at baseline. As noted in Step 3, we use data on the roughly 300 connected households, along with data on the 210 connected households in the baseline sample, mainly for descriptive purposes. Currently, we are administering a detailed follow-up survey of each household, capturing various measures of living standards ("Living Standards Kenya (LSK) Survey – Follow-up (2016)"). The follow-up survey round is expected to take place between May and October 2016.

1.3 Main and secondary samples

To summarize, our study will focus on two sets of households. The first set of households—which we refer to throughout this document as "main sample" households—consists of the 2,289 households that were unconnected to electricity at the time of the baseline survey. These households were randomly sampled using the baseline census data and are thus representative of the under grid population at baseline. Out of these 2,289 unconnected households, 1,139 were provided with opportunities to connect to the grid at a subsidized price, and 478 eventually chose to connect to the national grid. We have both baseline and follow-up survey data for the main sample households.

The second set of households—which we refer to as "secondary sample" households—will consist of the roughly 1,200 households that were observed to be unconnected at the time of the baseline census, but were not enrolled into the data collection during the baseline survey. These households were also randomly sampled using the baseline census data and are thus representative of the under grid population at baseline. Data from the secondary sample will allow us to study the spillover impacts of household electrification.

1.4 Analysis and data examined to date

At the time of registering this pre-analysis plan, we had collected follow-up survey information on over 3,500 households. Note that we did not perform any data analysis before registering this plan. As described in the document titled, "Note on data management/access and pre-analysis plan," which was uploaded to the AEA RCT Registry on May 16, 2016, the authors of this pre-analysis plan were provided with access to de-identified survey data for roughly 400 surveys, at the very beginning of the survey round. These data were stripped of any indicators that could expose the treatment status of households, and were provided in order to (1) allow the authors to identify and correct any coding errors in the survey instrument, (2) make improvements to the choice sets for multiple-choice questions, (3) identify and amend questions that were taking too much time to administer, (4) address any other technical issues with the survey instrument (for instance, with the SurveyCTO software coding), and (5) make any final additions to the survey instrument to address minor questions that came up. Each member of the research team agreed to follow the data management/access plan.

As a result of these early data quality checks, we learned that there were missing observations for a small number of variables. In order to address this issue, project field staff will revisit certain transformer communities at the end of the survey round to collect missing data. The analyses described in Section 2 will utilize the complete set of data. In the appendix, however, we will present additional robustness checks in which we drop all data that were re-collected at the end of the survey round.

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 this 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. Analysis

2.1 General notes

Randomly lowering the price of an electricity connection at the community-level by 29, 57, and 100 percent, resulted in increases in take-up of 6%, 22%, and 95%, over the baseline, respectively.[3] Take up in the low and medium subsidy treatment arms was relatively low. In our analysis, we will estimate both treatment-on-treated (TOT) and intention-to-treat (ITT) impacts of electrification. ITT estimates will be obtained from specifications in which various outcomes of interest are regressed on a set of binary variables indicating the treatment status of the community. TOT estimates will be obtained from two-stage least squares specifications in

[3] See Lee, Miguel, and Wolfram (2016b) for details.

which the household's electrification status, or the transformer community's electrification rate, is instrumented with the set of treatment indicators.

Throughout this document, we refer to our subject population as "households." In our setting, residential structures are typically located in compounds that can sometimes consist of multiple households. Our subject population consists of households that were considered to be the "main household" in the residential compound at the time of the baseline survey. To construct our sample, we randomly sampled compounds from each transformer community and enrolled the primary household in the compound. All other households in each compound are referred to as "minor households."

In the majority of our main sample analyses, we will focus on the family of the respondent that was interviewed at baseline, regardless of whether the family is still living in the same location at the time of the follow-up survey. For certain outcomes, however, we will focus on the family (if any) that is currently living at the physical location where the baseline survey took place. This will allow us to examine an additional set of questions including, for example, whether locations that were electrified are more likely to remain inhabited, compared to locations that were not electrified.

2.2 Main sample impacts

We will first analyze the main sample and test the hypothesis that households connected to the electricity grid enjoy higher levels of living standards, and analyze effects on other economic and social outcomes. Using main sample data, we will estimate ITT results using the following equation:

$$y_{ic} = \beta_0 + \beta_1 T_{Lc} + \beta_2 T_{Mc} + \beta_3 T_{Hc} + X_c'\Lambda + Z_{1ic}'\Gamma + \epsilon_{ic} \quad (1)$$

where y_{ic} represents the outcome of interest for main sample household i in community c, and T_{Lc}, T_{Mc}, and T_{Hc} are binary variables indicating whether community c was randomly assigned into the low-value, medium-value, and high-value subsidy treatment arms, respectively. Following Bruhn and McKenzie (2009), we include a vector of community-level characteristics, X_c, containing the variables used for stratification during randomization. In addition, we include Z_{1ic}, a vector of household-level characteristics. Further details on the components of the covariate vectors are presented in Section 2.7. The variables in Z_{1ic} will sometimes be used in analyses of treatment effect heterogeneity, which is discussed in further detail in Section 2.8. In Section 2.10, we discuss the possibility of ANCOVA specifications for certain outcome variables. Standard errors will be clustered at the community level.

The issue of limited statistical power may be more severe in ITT specifications due to the relatively low take-up rates in the low and medium subsidy treatment groups. To address this issue, we will focus attention on the coefficient on the high subsidy treatment indicator. This test will not only shed light on the impacts of near universal electrification, but also is likely to have greater statistical power. The coefficients on the low and medium subsidy treatment indicators will also be of interest, since these will shed light on the average impacts of an electrification program with low take-up.[4]

We will also estimate TOT results using the following equations:

$$\mathrm{E_{ic}} = \delta_0 + \delta_1 \mathrm{T_{Lc}} + \delta_2 \mathrm{T_{Mc}} + \delta_3 \mathrm{T_{Hc}} + \mathrm{X}_c' \Lambda_1 + \mathrm{Z}_{1ic}' \Gamma_1 + \eta_{ic} \tag{2}$$

$$\mathrm{y_{ic}} = \beta_0 + \beta_1 \mathrm{E_{ic}} + \mathrm{X}_c' \Lambda_2 + \mathrm{Z}_{1ic}' \Gamma_2 + \epsilon_{ic} \tag{3}$$

where the first-stage equation 2 estimates the effects of the treatment indicators on household electrification status, $\mathrm{E_{ic}}$, and the second-stage equation 3 estimates the effect of household electrification status on the various outcomes of interest. As in equation 1, errors will be clustered at the community level.

Lee, Miguel, and Wolfram (2016b) document systematic differences in the baseline living conditions of households taking up the experimental offers in the low and medium subsidy groups, compared to the high subsidy group. Households that paid more for an electricity connection (i.e., low subsidy arm households) were wealthier and more educated on average than those who paid nothing (i.e., high subsidy arm households). This suggests that the average treatment effect may vary across treatment arms. For example, electrification may be more impactful for the relatively wealthier households that are able to invest in complementary assets such as electrical appliances. In order to examine these types of heterogeneous treatment effects, we may explore the methods described in Kowalski (2016) to first recover bounds on average treatment effects for "always taker" and "never taker" households, and then decompose group average treated outcomes into selection and treatment effects. However, due to relatively

[4] Note that the effective price of an electricity connection in the medium subsidy arm is $171 (or 15,000 KSh), which is the same price as that offered under the World Bank and African Development Bank-funded Kenya Last Mile Connectivity Project. These estimated effects are therefore likely to be of policy interest.

low take-up rates in the low and medium subsidy groups, these analyses may be statistically underpowered.

Although the experiment generated exogenous variation in household electrification status, there remain some challenges in econometric identification. For example, if there are substantial local spillovers for unconnected or connected households, the stable unit treatment value assumption (SUTVA) may not hold. In this case, it is methodologically preferable to focus on the ITT results, and in particular, the coefficient on the high subsidy treatment indicator since it has a clearer interpretation. We describe our plan to quantify spillovers in the next section.

2.3 Community-level outcomes

For community-level outcomes (which are specified in Section 3.12), we will estimate equations that are similar in form to those specified in Section 2.2, with the exception of two key differences. First, we will use both main and secondary sample data to construct the community-level outcomes of interest. Second, since the unit of observation is the community, we will exclude household-level covariates.

In the TOT specification to estimate community-level impacts, we will replace the E_{ic} term in equations 2 and 3 with R_c, the estimated local transformer community electrification rate, which itself is a major outcome of interest. Note that for each transformer community, we have data on the universe of households (as well as their grid connection status) at the time of our baseline census. In addition, we have follow-up household survey data for the main and secondary sample households. Since we do not have updated census data for each transformer community, we will need to estimate the current rate. For each of the three treatment arms, we will calculate the average take-up rate for the portion of secondary sample households that were observed to be unconnected at the time of the baseline census. We will estimate R_c by combining actual follow-up take-up data among the surveyed households with estimated take-up data for the non-surveyed households (i.e., those observed to be unconnected at the time of the baseline census) in the relevant treatment group. Specifically, for each treatment arm, we will assume that all of the remaining, non-surveyed households connected to the grid at the treatment arm-level average take-up rate. For the control group communities, we will also include main sample households when calculating the control group take-up rate that will then be applied to the non-surveyed control group households. See Section 3.12 for additional details on how we plan to construct community-level outcome variables.

2.4 Secondary sample impacts

We consider two types of potential spillovers. First, as shown in Bernard and Torero (2015), it is possible that an exogenous increase in the local electrification rate will encourage neighboring unconnected households to connect to the grid. In this case, we would expect to find higher electrification rates—as well as higher levels of willingness to pay for electricity—among secondary sample households in treatment communities, compared to control communities. We discuss our planned analysis of willingness to pay in Section 2.6. Second, it is possible that private grid connections result in economic and social impacts for neighboring households, for instance, if they sometimes use their neighbors' power. In this case, we would expect to find improved living standards for secondary sample households located in treatment communities.

Using secondary sample data, we will estimate ITT results using the following equation:

$$y_{ic} = \beta_0 + \beta_1 T_{Lc} + \beta_2 T_{Mc} + \beta_3 T_{Hc} + X_c' \Lambda + Z_{2ic}' \Gamma + \epsilon_{ic} \qquad (4)$$

where Z_{2ic} is the vector of household characteristics (see Section 2.7). We differentiate between Z_{1ic} and Z_{2ic} to account for a few covariates that are specific to either the main or secondary sample households. In order to concentrate attention on a coefficient with sufficient statistical power, we will again focus on the coefficient on the high subsidy treatment indicator, T_{Hc}. Recall that secondary sample households in the high subsidy treatment arm are likely to have a far higher number of recently connected neighbors.

We will also estimate TOT results for the secondary sample, but will take a slightly different analytical approach. First, we will estimate the equations:

$$R_c = \delta_0 + \delta_1 T_{Lc} + \delta_2 T_{Mc} + \delta_3 T_{Hc} + X_c' \Lambda_1 + Z_{2ic}' \Gamma_1 + \eta_{ic} \qquad (5)$$

$$y_{ic} = \beta_0 + \beta_1 R_c + X_c' \Lambda_2 + Z_{2ic}' \Gamma_2 + \epsilon_{ic} \qquad (6)$$

where R_c is the estimated local transformer community electrification rate (described in Section 2.3 above). The first-stage equation 5 estimates the effects of the treatment variables on the community electrification rate. The second-stage equation 6 estimates the effect of the community electrification rate on the household-level outcomes of interest. Second, we will estimate the set of equations:

$$\begin{aligned} R_c = \; & \delta_{0,R} + \delta_{1,R}T_{Lc} + \delta_{2,R}T_{Mc} + \delta_{3,R}T_{Hc} + \omega_{0,R}d_{ic} + \omega_{1,R}(T_{Lc}\times d_{ic}) + \\ & \omega_{2,R}(T_{Mc}\times d_{ic}) + \omega_{3,R}(T_{Hc}\times d_{ic}) + X_c'\Lambda_R + Z_{2ic}'\Gamma_R + \eta_{ic,R} \qquad (7) \\ r_{ic} = \; & \delta_{0,r} + \delta_{1,r}T_{Lc} + \delta_{2,r}T_{Mc} + \delta_{3,r}T_{Hc} + \omega_{0,r}d_{ic} + \omega_{1,r}(T_{Lc}\times d_{ic}) + \\ & \omega_{2,r}(T_{Mc}\times d_{ic}) + \omega_{3,r}(T_{Hc}\times d_{ic}) + X_c'\Lambda_r + Z_{2ic}'\Gamma_r + \eta_{ic,r} \qquad (8) \\ y_{ic} = \; & \beta_0 + \beta_1 R_c + \beta_2 r_{ic} + \omega_{0,2}d_{ic} + X_c'\Lambda_2 + Z_{2ic}'\Gamma_2 + \epsilon_{ic} \qquad (9) \end{aligned}$$

where equations 7 and 8 are the first-stage equations and equation 9 is the second-stage equation. R_c is again defined as the estimated local transformer community electrification rate, r_{ic} represents the proportion of households within 200 meters of household i that are connected to electricity, and d_{ic} represents the proportion of households within 200 meters of household i that are in the main sample. We instrument r_{ic} with the treatment indicators, T_{Lc}, T_{Mc}, and T_{Hc}, as well as d_{ic} and the interactions between d_{ic} and the treatment indicators. The second-stage equation 9 therefore estimates the effects of the community electrification rate, as well as being in close proximity to connected households, on the outcome y_{ic}. Since there are multiple endogenous variables in this framework, equations (7), (8), and (9) will be estimated jointly.

The secondary sample analysis will allow us to determine whether there are any meaningful spillovers from household grid connections. This will in turn guide our interpretation of the main sample analysis in the ways noted above, especially in relation to the validity of the proposed instrumental variables approach. Note that it is challenging to precisely define the exact pattern of results that will allow us to conclude that the spillovers are "meaningful". Broadly, if we estimate statistically significant spillover impacts on a number of key outcomes, then we will mainly focus on the main sample analysis ITT specifications rather than the TOT specifications in Section 2.2. Similarly, if we estimate statistically significant impacts on the connection rates of secondary sample households, the proposed instruments described in equations 5 through 9 would also violate the exclusion restriction and it may be preferable to focus on the ITT specification in equation 4.

2.5 Educational impacts

Another objective of this study is to understand the extent to which household electrification impacts the educational outcomes of schoolchildren. As part of the follow-up survey round, we administered short (roughly 15 minute) reading and math tests to all 12 to 15 year olds in our subject population. Using these data, we will estimate regressions that are similar in form to those specified in Sections 2.2 and 2.4 but will focus on individual children as the unit of observation. In these regressions, the covariate vectors $\mathrm{Z_{1ic}}$ (for the main sample) and $\mathrm{Z_{2ic}}$ (for the secondary sample) will be complemented with the covariate vector $\mathrm{C_{jic}}$, which includes additional information on child j in household i in community c. The outcomes of interest in these specifications will therefore be denoted with the subscript jic. The covariate vector $\mathrm{C_{jic}}$ is described in more detail in Section 2.7.

2.6 Stated willingness to pay (WTP) for electricity

In the follow-up survey, we first ask respondents whether they would be hypothetically willing to connect to the national grid at a randomly selected price (i.e., *time unlimited* offer) (f3g in the follow-up survey). The randomly selected price, p, was drawn from the following set of prices (in Kenyan shillings):

$$\{0, 5000, 10000, 15000, 20000, 25000, 35000, 75000\}$$

This question was followed by an additional hypothetical question (f3h) asking the respondent whether they would accept an offer at this price if they were given only six weeks to complete the payment (i.e., *time limited* offer). Finally, respondents were asked whether they would be willing to pay a monthly amount over a period of three years, where the cumulative total is equal to the randomly selected price (f3i) (i.e., *financed offer*, with terms similar to those offered under the current Kenya LMCP). Respondents from connected households were asked a similar set of questions with somewhat different wording to reflect the fact that they are already connected (see f4g, f4h, and f4i).

We are interested in understanding how stated WTP responds to price levels. Specifically, we will estimate the following equation:

$$\begin{aligned}\mathrm{h_{ic}} = {} & \alpha_0 + \alpha_1 \mathrm{T_{Lc}} + \alpha_2 \mathrm{T_{Mc}} + \alpha_3 \mathrm{T_{Hc}} + \textstyle\sum_{\mathrm{p}} \beta_{\mathrm{p}} \mathrm{W_{pic}} + \textstyle\sum_{\mathrm{p}} \gamma_{1\mathrm{p}} (\mathrm{W_{pic}} \times \mathrm{T_{Lc}}) + \\ & + \textstyle\sum_{\mathrm{p}} \gamma_{2\mathrm{p}} (\mathrm{W_{pic}} \times \mathrm{T_{Mc}}) + \textstyle\sum_{\mathrm{p}} \gamma_{3\mathrm{p}} (\mathrm{W_{pic}} \times \mathrm{T_{Hc}}) + \mathrm{X'_c}\Lambda + \mathrm{Z_{nic}}\Gamma + \epsilon_{\mathrm{ic}} \qquad (10)\end{aligned}$$

where h_{ic} is a binary variable indicating the stated (or hypothetical) take-up decision for household i, W_{pi} is a binary variable indicating whether household i received the hypothetical price p, and Z_{nic} is the relevant household covariate vector. We are especially interested both in the direct effects of the treatment indicators, as well as the coefficients on the full set of interactions between the treatment indicators and the W_{pic} terms. These interactions will shed light on how stated WTP may be different for households that were recently connected to the grid (e.g., using the main sample data), or for unconnected households that recently observed neighboring households become connected to the grid (e.g., using the secondary sample data).

We will estimate separate regressions for the main sample and the secondary sample, since the interpretation of the results will be slightly different for each case. Standard errors will be clustered at the community level.[5] We will also test for heterogeneous effects, which are generally described in Section 2.8.

As in Lee, Miguel, and Wolfram (2016b), we will plot the stated WTP results graphically. For example, we may plot and compare demand curves for (1) time unlimited, time limited, and financed offers, (2) control households at baseline and at follow-up, (3) main sample households in the various subsidy arms and in the control group at follow-up, and (4) secondary sample households in the various subsidy arms and in the control group, as well as other leading comparisons.

2.7 Covariate vectors X_c, Z_{1ic}, Z_{2ic}, and C_{jic}

In this section, we describe each of the sets of covariates that we plan to utilize in the analysis.

The vector X_c will primarily include the stratification variables that were used during randomization. These include:

- County: Binary variable indicating whether community c is in Busia county or Siaya county.
- Market status: Binary variable indicating whether the total number of businesses in community c is strictly greater than the community-level mean across the entire sample.

[5] Based on the results of Lee, Miguel, and Wolfram (2016b), we do not expect the relationship between take-up and price to be linear. However, we may still test for linearity, and if we cannot reject linearity in an F-test, we will also estimate an equation in which y_{ic} is regressed on p_{ic}, controlling for the treatment indicators and other covariates.

We use this definition to define which communities could be classified as "markets" relative to others.

- Transformer funding year: Binary variable indicating whether the electricity transformer in community c was funded "early" (i.e. in either 2008-09 or 2009-10).
- Electrification rate: Residential electrification rate in community c at the time of census (roughly 2013).
- Community population: Estimated number of people living in community c at the time of census (roughly 2013).

The vector Z_{1ic}, which will be included in regressions using the main sample data, will include the set of household-level variables listed below. Note that for the main sample households, we will be able to take advantage of the baseline survey data.

- Gender of respondent: Binary variable indicating whether the respondent is female.
- Age: Age of respondent in 2016.
- Education of respondent at baseline: Binary variable indicating whether the household respondent at baseline has completed secondary school.[6]
- Bank account at baseline: Binary variable indicating whether the household respondent at baseline had a bank account.
- Housing quality index at baseline: Index composed of whether the household had high-quality floors, roof, and walls at baseline.
- Asset value at baseline: Estimated value based on inventory of livestock, electrical appliances, and non-livestock assets at baseline, at current observed local prices.
- Energy spending at baseline: Estimated monthly expenditures on all energy sources at baseline.

The vector Z_{2ic}, which will be included in regressions using the secondary sample data, will include the household-level variables listed below. Note that there is no baseline survey data for this sample of households.

- Gender of respondent: Binary variable indicating whether the respondent is female.
- Age: Age of respondent in 2016.

[6] The respondent during the baseline survey is not necessarily the same person as the respondent during the follow-up survey.

- Local density: Total number of households in the transformer community within 200 meters.[7]

The vector C_{jic} will include a set of individual-level characteristics that are relevant for the regression specifications estimating the impacts of electrification on educational performance.

- Gender of student: Binary variable indicating whether the student is female.
- Age: Age of student in 2016.
- Siblings: Number of children under 18 in the household.
- Grade attained at baseline (main sample only): Grade attained by the end of the 2013 academic year.[8]

2.8 Heterogeneous effects

In additional analyses, we will estimate heterogeneous treatment effects along a number of major dimensions, captured in the vectors X_c, X_{1ic}, X_{2ic}, and C_{jic}, by adding interaction terms between each treatment indicator and these variables. For instance, in order to assess how treatment impacts may vary for households at different wealth levels, we will estimate specifications in which the treatment indicators are interacted with the housing quality index at baseline.

Furthermore, there are a number of additional (and potentially endogenous) variables that are not included in the covariate vectors above but are of potential interest. These include:

- Transformer outages in the community: Proportion of months (between September 2014 and October 2015) that the transformer was not working.
- Connection days: Approximate number of days since the household was first connected to electricity.
- Relationships with main sample households (for secondary sample households): Number of main sample households whose members are considered to be extended family of the secondary sample respondent.

[7] In additional robustness checks, we will also carry out analysis using the total number of households in the transformer community within 400 meters.

[8] We will infer this data by comparing the baseline and follow-up surveys for main sample households. It is possible that this data will be missing for a large number of observations. In these instances, we may include an additional binary variable indicating that the data are missing. Alternatively, we may choose to drop this covariate altogether if this data are missing for over 30% of possible values from collected surveys.

We are uncertain whether our study design will have sufficient statistical power to generate precise estimates on many of these interaction terms and hence such analyses should be considered suggestive rather than definitive. The patterns that emerge will also likely stimulate further exploratory analysis using the dataset.

2.9 Construction of indices

When constructing indices, we will normalize each component variable to have mean zero and unit variance, thereafter constructing the index by summing each component variable (the mean effects approach). Note that we will exclude any variables with zero variance since these do not contribute any information to the analysis. Furthermore, if a pre-specified variable is missing more than 30% of possible values from collected follow-up surveys, we will drop it from inclusion in 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. Finally, we will report all individual outcomes used to create indices in the appendix.

2.10 Multiple Testing Adjustment

In Section 3, we describe how the major outcomes of interest are categorized into ten broad "families". For the main coefficient estimates of interest (for instance, β_1, β_2, and β_3 in equation 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. For instance, researchers interested in the effect of household electrification on non-agricultural compensation should focus directly on this p-value.

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 null hypothesis.[9] Specifically, we will follow the approach to FDR analysis adopted in Casey et al. (2012) and the references cited therein (e.g., Anderson 2008).

2.11 Additional analyses

For a subset of outcomes in the main sample regressions, we will have both baseline and follow-up observations (e.g., household size, home solar system usage, energy consumption, etc.). In this case, we will also estimate ANCOVA regression specifications in which the baseline value of the outcome of interest is included as an additional covariate, as the resulting estimates may have greater statistical power (McKenzie 2012). However, note that we lack equivalent baseline measures for many outcome variables described below (in Section 3). This is particularly the case when the household respondent in the follow-up survey is not the same person as the household respondent in the baseline survey. As a result, the ANCOVA estimates will be presented mostly as a supplement. Our main focus will be on the results of the specifications described in Sections 2.2 and 2.4 above.

3. Major outcomes of interest

3.1 Overview

In this section, we specify 77 major economic and social outcomes of interest. These outcomes have been selected based on the judgment of the research team and are arranged into ten broad families: (1) energy consumption, (2) household structure, (3) time use, (4) productivity, (5) wealth, (6) consumption, (7) health and wellbeing, (8) education, (9) social and political attitudes, and (10) community outcomes. Based on this list, we also identify a group of ten "primary" outcomes, drawn from a number of different outcome families. The estimated impacts on these primary outcomes will serve as an overall summary of the impacts of household electrification in our setting. As discussed in Section 2.10, we will present FDR-adjusted q-values for each of the outcomes within the primary outcomes group, as well as FDR-adjusted q-values for each outcome within each of the ten outcome families. As noted in Section 1.4, we anticipate that we will examine additional outcomes beyond those included in this plan.

[9] 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.

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

Table 1 summarizes the ten primary outcomes that will serve as an overall summary of living standards in our setting.

Table 1. Primary outcomes

ID	Outcome	Unit	Type	Description	Ref.
P.1	Grid connected	HH	Indicator	Indicator for main household connection	1.1
P.2	Grid electricity spending	HH	Total	Estimated prepaid top-up last month or amount of last postpaid bill	1.7
P.3	Employed or own business - Household	HH	Proportion	Proportion of household members (18 and over) currently employed or running their own business	4.5
P.4	Total hours worked	Resp.	Total	Total hours worked in agriculture, self-employment, employment, and household chores in last 7 days	4.11
P.5	Total asset value	HH	Estimated value	Estimated value of savings, livestock, electrical appliances, and other assets	5.6
P.6	Annual consumption	HH	Value	Estimated value of annual consumption of 23 goods	6.2
P.7	Recent symptoms index	Resp.	Index	Index of symptoms experienced by the respondent over the past 4 weeks	7.3
P.8	Life satisfaction	Resp.	Scale	Life satisfaction based on a scale of 1 to 10	7.8
P.9	Average test score	Child	Z-score	Average of English reading test result and Math test result	8.3
P.10	Political and social awareness index	Resp.	Index	Index capturing the extent to which the respondent correctly answered a series of questions about current events	9.4

For certain primary outcomes, we are able to use the existing literature to guide our expectations on the impacts of electrification in our setting. For example, in South Africa, Dinkelman (2011) finds that female employment rises by 9 to 9.5 percentage points and women work roughly 8.9 hours more per week. In Brazil, Lipscomb, Mobarak, and Barham (2013) find that the probability of employment increases by 17 to 18 percentage points, over the long run, in counties that are electrified. Taken together, we should expect to find substantial increases in the probability of employment (P.3) and labor hours (P.4), particularly for women. Furthermore, in the Philippines, Chakravorty, Emerick, and Ravago (2016) estimate that village-level electrification leads to an increase in household expenditures by 38 percent, suggesting that there will be large gains in household consumption (P.6). In terms of test scores (P.9), Hassan and Lucchino (2016) examine the impacts of randomly distributing solar lanterns to 7^{th} grade pupils in Kenya and find math grades to increase by 0.88 standard deviations for treatment pupils. In our analysis of each primary outcome, we will test the null hypothesis and (wherever possible) the hypothesis that the treatment effect is the same as that found in the existing literature. Finally, we will compare the estimated impacts in our study to other outcomes in the broader development economics literature in order to assess the cost effectiveness of rural electrification as a development policy.

3.3 Family #1 – Energy consumption major outcomes

At the most basic level, electricity connections should impact the way in which households consume energy. Family 1 includes the major outcomes relating to access to and usage of different forms of energy.

Table 2. Energy consumption major outcomes

ID	Outcome	Unit	Type	Component(s)	Survey data
1.1	Grid connected	HH	Indicator	Indicator for main household connection	F1a
1.2	Electric lighting	HH	Indicator	Indicator for electricity as main source of lighting	F1b
1.3	Lighting usage	HH	Total	Hours of lighting used (past 24 hours)	F18
1.4	Installation	HH	Total	Number of electrical outlets available Number of lighting sockets available Number of power strips in use	F6b F6c F6e
1.5	Appliances owned	HH	Total	Number of “high-wattage” appliances owned[10]	F19a to F19c
1.6	Appliances desired	HH	Total	Number of “high-wattage” appliances desired	F19d to F19g

[10] In general, we follow Lee, Miguel, and Wolfram (2016a) in the definition of high and low wattage appliances. For instance, there we define mobile phones and radios as “low-wattage” appliances.

1.7	Grid electricity spending	HH	Total	Estimated prepaid top-up last month Amount of last postpaid bill	F7a to F7e, F5h F8a to F8c, F5h
1.8	Kerosene spending	HH	Total	Kerosene spending last month[11]	F11
1.9	Other energy sources spending[12]	HH	Total	Solar power spending last month Battery spending last month Generator spending last month Purchased firewood spending last month Charcoal spending last month LPG spending last month Sawdust spending last month Mobile phone charging last month Other spending last month	F13d, F14d F15b, F15c F16c F17a F17b F17c F17d F17h F17e to F17g, F17i
1.10	Total energy spending	HH	Total	Total spending last month on grid electricity, kerosene, and other energy sources	See 1.7, 1.8, and 1.9 above
1.11	Home solar usage	HH	Indicator	Indicator for usage of solar lantern or solar home system	F12a
1.12	Power sharing	HH	Indicator	Indicator for household is sharing its electricity connection (e.g., electricity connection shared with a minor household or a neighboring household)	S1c, F5b, F5i, F5j

3.4 Family #2 – Household structure major outcomes

If there are changes in the patterns of energy consumption, there may also be changes in the structure of the household. For example, access to electricity may impact household structure by influencing incentives to migrate by making living in the household more attractive. Family 2 includes major outcomes relating to household structure, migration, and fertility.

Table 3. Household structure major outcomes

ID	Outcome	Unit	Type	Component(s)	Survey data
2.1	Household size	HH	Total	Total number of household members	Section A, hhsize
2.2	Inhabited location	HH	Indicator	Baseline structure currently inhabited	Staff records
2.3	Household stayed	HH	Indicator	Household did not move to a new location	Staff records, AA9
2.4	Members living elsewhere	HH	Total	Household members documented at baseline that are now living elsewhere	Section A

[11] For several energy spending categories (including kerosene), we recorded how much the household spent over the past seven days. In these cases, we will estimate spending over the past month by multiplying the weekly amount by a factor of approximately 4.3.

[12] This outcome will include all other energy-related expenditures recorded in the household survey, beyond grid electricity and kerosene.

2.5	Fertility	Resp.	Total	Number of times respondent (or sexual partner) has been pregnant since January 2014	sH3_3num_m, sH3_3num_f
2.6	Local social interactions	Resp.	Total	Number of times (over past week) neighboring respondents visited household and respondent visited neighboring households	Section K

3.5 Family #3 – Time use major outcomes

Household electrification may operate as a labor saving technology shock to home production, releasing female time from home to market work (Dinkelman 2011; Grogan and Sadanand 2012). Family 3 includes individual time use outcomes.

Table 4. Time use major outcomes

ID	Outcome	Unit	Type	Component(s)	Survey data
3.1	Hours sleeping	Resp.	Hours	Sleeping (code 1)	L1 to L48
3.2	Hours studying	Resp.	Hours	Playing with children or helping with homework (code 13) Studying or attending class (code 16) *Note: All codes representing "studying" in survey*	L1 to L48
3.3	Hours working	Resp.	Hours	Light farm work (code 22) Heavy farm work (code 23) Fishing or hunting (code 24) Office/desk work (code 25) Light manual work (code 26) Heavy manual work (code 27) Other (work and travel) (code 32) *Note: All codes representing "work" in survey*	L1 to L48
3.4	Hours doing chores	Resp.	Hours	Cooking or preparing food (code 7) Shopping for family (code 8) Cleaning, dusting, sweeping, washing dishes or clothes, ironing, or doing other household chores (code 9) Taking care of others, such as bathing, feeding, or looking after children, the sick, or the elderly (code 12) Fetching water or firewood (code 10) Repairs in or around the home (code 11) Improving land or buildings (code 28) *Note: All codes representing "chores" in survey*	L1 to L48
3.5	Hours enjoying leisure	Resp.	Hours	Rest, watching TV, listening to the radio, reading a book, watching a movie, watching sports, or sewing (code 6) Visiting or entertaining friends (code 14) Playing sports (code 17) Spending time with spouse or partner (code 18) *Note: All codes representing "leisure" in survey*	L1 to L48

3.6 Family #4 – Productivity major outcomes

If electrification changes people's time use, and, for example, allows for more hours of work outside the home, there may be positive impacts on various measures of productivity and wealth.[13] The evidence on the impacts of electrification on productivity have been somewhat mixed. Dinkelman (2011), for example, finds evidence of increased female labor force participation in South Africa. Chakravorty, Emerick, and Ravago (2016) find large impacts of electrification on household income and expenditures in the Philippines, but attribute these impacts to increases in agricultural income rather than increases in labor force participation. In contrast, Burlig and Preonas (2016) find little to no impacts of electrification on various employment outcomes in rural India. Family 4 includes various measures of household agricultural activities, employment, small businesses, and other outcomes.

Table 5. Productivity major outcomes

ID	Outcome	Unit	Type	Component(s)	Survey data
4.1	Agriculture – Land use	HH	Proportion	Proportion of total land used for agricultural activities	C4a, C4b, D1c
4.2	Irrigation	HH	Indicator	Household used irrigation in last 12 months	D2e
4.3	Agriculture – Monthly revenue	HH	Total	Revenue from selling crops	D4a
				Revenue from selling livestock or livestock products	D4c
				Revenue from selling poultry or poultry products	D4e
				Revenue from selling fish	D4g
				Revenue from selling other agricultural produce *Note: Household revenue over past month*	D4i
4.4	Agriculture – Hours worked	Resp.	Total	Hours worked in agriculture in last 7 days	D3a
4.5	Employed or own business - Household	HH	Proportion	Proportion of household members (18 and over) currently employed or running their own business	A8b
4.6	Business at household	HH	Indicator	Business operated out of household compound	sE1_15cdescpremise, sE1_51otherbus
4.7	Employed or own business – Individual	Resp.	Indicator	Currently self-employed, running a business, employed, or working for pay	sE1_1selfemp, sE2_1employed
4.8	Employed or own business – Individual monthly compensation	Resp.	Total	Monthly compensation, sum of last month compensation across all jobs and businesses	sE2_11, sE1_9aprofit, sE1_56profit

13 Grimm et al. (2015), for instance, present a theoretical model in which an increase in household electrification effectively reduces the price of energy faced by the household, which increases the productivity of domestic labor and the output of household production.

4.9	Employed or own business – Individual hours worked	Resp.	Total	Hours worked in self-employment in last 7 days Hours worked in employment in last 7 days	sE1_5wrkhrs sE2_7hours_1
4.10	Household chores – Individual hours worked	Resp.	Total	Hours spent doing household chores in last 7 days	sL_49hhchores
4.11	Total hours worked	Resp.	Total	Total hours worked in agriculture, self-employment, employment, and household chores in last 7 days	See 4.3, 4.8, and 4.9 above

3.7 Family #5 – Wealth major outcomes

In terms of wealth, Lipscomb, Mobarak, and Barham (2013) find evidence of higher average housing values as a result of electrification in Brazil. Family 5 includes a housing quality index and estimated values of different types of household assets, based on current market prices.

Table 6. Wealth major outcomes

ID	Outcome	Unit	Type	Component(s)	Survey data
5.1	Savings	Resp.	Total	Savings in mobile bank account Savings in SACCO, merry-go-round, or ROSCA Savings in formal bank account	G2a G2b G2c
5.2	Housing quality	HH	Index	Indicator for high-quality floors Indicator for high-quality roof Indicator for high-quality walls	C1a C1b C1c
5.3	Value of livestock assets	HH	Estimated value	Value of chickens owned Value of cattle owned Value of goats owned Value of pigs owned Value of sheep owned	C8a C8b C8c C8d C8e
5.4	Value of appliance assets	HH	Estimated value	Value of listed electrical appliances	F19a to F19c
5.5	Value of other assets	HH	Estimated value	Value of beds owned Value of bednets owned Value of kerosene stoves owned Value of kerosene lamps owned Value of hoes owned Value of bicycles owned Value of motorcycles owned Value of cars or trucks owned Value of sofa piece seats owned	C7a C7b C7c C7d C7e C7f C7g C7h C7i

5.6	Total asset value	HH	Estimated value	Estimated value of savings, livestock, electrical appliances, and other assets	See 5.1, 5.3, 5.4, and 5.5 above

3.8 Family #6 – Consumption major outcomes

We are interested in estimating the impacts of electrification on various measures of household consumption, including a novel "neediness" index, developed in Ligon (2015). The neediness index is a measure of the marginal utility of expenditures and therefore household welfare. Unlike traditional total consumption expenditure measures, it does not impose an assumption of linear Engel curves. Instead, the index exploits differences in the composition of consumers' consumption bundles, which vary with household welfare. In order to construct the index, Ligon (2015) suggests collecting information on a subset of key consumption items for which variation in expenditures is closely related to changes in marginal utility (and thus welfare). By appropriately weighting the consumption of each of the key items, we can obtain a summary measure of household welfare. In our setting in Western Kenya, we will focus on 23 items, including staples, vegetables, meat, fruits, and other goods. These 23 items were identified using data from the Kenya Life Panel Survey (KLPS-3).[14] Based on the KLPS-3 data, the 23 items account for 26% of total household consumption and 52% of total food consumption.

Table 7. Consumption major outcomes

ID	Outcome	Unit	Type	Component(s)	Survey data
6.1	Neediness index	HH	Index	Consumption of each of 23 goods over past twelve months, constructed according to the measure in Ligon (2015)	M5, M7, M8
6.2	Annual consumption	HH	Value	Estimated value of annual consumption of 23 goods	M5, M7, M8
6.3	Consumption diversity	HH	Index	Indicators for whether household has consumed each of 23 goods over the past twelve months	M1
6.4	Meals	Resp.	Total	Total number of meals eaten yesterday	sH1_1meals
6.5	Protein meals	Resp.	Total	Total number of meals eaten yesterday including meat or fish	sH1_2ameat

3.9 Family #7 – Health and wellbeing outcomes

Electricity has been found to improve respiratory health by reducing indoor air pollution (Barron and Torero 2015). Some people may also be happier when they have access to

[14] The KLPS-3 project is located in the same study region as this project and is led by PI Edward Miguel and other researchers. In the full KLPS survey, respondents are asked in detail about their consumption of 153 items.

electricity due to impacts on various channels. Family 7 includes various measures of respondent health and wellbeing.

Table 8. Health and wellbeing major outcomes

ID	Outcome	Unit	Type	Component(s)	Survey data
7.1	Respiratory illness index	Resp.	Index	Persistent cough Asthma/breathlessness at night *Note: Experienced over past 4 weeks*	sH1_7bcough sH1_7sasthma
7.2	Respiratory illness index - Child	Child	Index	Frequent cough Itchy or stinging eyes Sore throat Runny nose Asthma or breathlessness *Note: Experienced over past 7 days*	T3.5
7.3	Recent symptoms index	Resp.	Index	Fever Persistent cough Persistent tiredness Stomach pain Blood in stool Rapid weight loss Frequent diarrhea Skin rash or irritation Open sores/boils Difficulty swallowing Sores or ulcers on the genitals Asthma/breathlessness at night Frequent and excessive urination Constant thirst/increased drinking of fluids Unusual discharge from the tip of penis *(for men only)* Other symptoms *Note: All symptoms experienced over past 4 weeks*	sH1_7afever sH1_7bcough sH1_7ctired sH1_7dstomach sH1_7fstool sH1_7gweightloss sH1_7hdiarrhoea sH1_7iskin sH1_7jboils sH1_7kswallow sH1_7pgenitalsore sH1_7sasthma sH1_7tfrequrine sH1_7uthirst sH1_7wdischarge sH1_7xother
7.4	Recent illnesses index	Resp.	Index	Worms Malaria Typhoid Tuberculosis Diabetes Cholera Yellow fever *Note: All illnesses experienced over past 4 weeks*	sH1_7eworms sH1_7mmalaria sH1_7ntyphoid sH1_7otb sH1_7vdiabetes sH1_7qcholera sH1_7ryellow
7.5	Recent illnesses index - Child	Child	Index	Malaria Fever	T3.5

				Typhoid *Note: All symptoms experienced over past 7 days*	
7.6	Subjective health	Resp.	Indicator	Self-described health is either "good" or "very good"	sH1_13healthgd
7.7	Subjective health - Child	Child	Indicator	Self-described health is either "good" or "very good"	T3.4
7.8	Life satisfaction	Resp.	Scale	Life satisfaction based on a scale of 1 to 10	J9b

3.10 Family #8 – Education outcomes

It is possible that electrification may improve educational outcomes for students, if better lighting allows for more evening study time, for instance. The evidence, however, has been somewhat mixed to date. Randomized trials, including Furukawa (2014) and Hassan and Lucchino (2016), have focused on measuring the impacts of decentralized power solutions, such as solar lanterns, and have documented results ranging from negative impacts to positive impacts with substantial spillovers. Studies on the impacts of grid connections have been mostly non-experimental and have found positive impacts of electrification on school enrollment, study time, and school completion (see, e.g., Khandker et al. 2012). Family 8 includes a variety of educational outcomes, including test scores from English and Math tests that were administered to students in the sample villages by our project field staff.

Table 9. Education major outcomes

ID	Outcome	Unit	Type	Component(s)	Survey data
8.1	English score	Child	Z-score[15]	English reading test result	T1
8.2	Math score	Child	Z-score	Math test result	T2
8.3	Average test score	Child	Z-score	Average of English reading test result and Math test result	T1, T2
8.4	Study hours - Total	Child	Total	Self-reported hours spent studying during the day Self-reported hours spent studying during the night	T3.1 T3.2
8.5	Study hours - Night	Child	Total	Self-reported hours spent studying during the night	T3.2
8.6	Attendance index	Child	Index	Fully completed first week of school last term Fully completed last week of school last term Completed end of term exams last term Fully completed first week of school this term	B2b B2c B2d B2e
8.7	Grades	Child	Score	Marks (scaled out of 100) earned last term	B2f
8.8	Ambitions	Child	Indicator	Student planning to attend post-secondary education	T3.7

[15] We will create Z-scores by subtracting the mean and dividing by the standard deviation in the control group, within our own sample using age-gender groups.

3.11 Family #9 – Social and political attitudes outcomes

Electrified households may consume more media content (via televisions, radios, and internet access), and as a result, could have greater knowledge of current affairs, or experience changes in social and political attitudes.

Table 10. Social and political attitudes major outcomes

ID	Outcome	Unit	Type	Details	Question
9.1	Radio	Resp.	Total	Days in the past week respondent listened to the radio	J2a
9.2	Television	Resp.	Total	Days in the past week respondent watched television	J2c
9.3	Internet	Resp.	Total	Days in the past week respondent used the internet	J2d
9.4	Political and social awareness index	Resp.	Index	Knows date of next election	J1a
				Knows name of the president of Tanzania	J1b
				Knows name of the president of Burundi	J1c
				Knows name of a candidate in the 2016 U.S. presidential election	J1d
				Knows name of the CEO of Safaricom	J1e
				Knows name of the Managing Director of Kenya Power	J1f
				Knows the intended recipients of the Kenyan national government's Free Laptop program	J1i
				Knows who was responsible for the 2015 terrorist attacks at Garissa University	J1j
				Knows which team won the 2015-2016 English Premier League	J1g
				Knows who sings the pop song "Sura Yako" *Note: These are all binary variables*	J1h
9.5	Approval of national government index	Resp.	Index	Trusts national government	J5g
				Uhuru Kenyatta is doing a good job as president	J7a
				Government is doing a good job fighting terrorism	J7b
				Government corruption is <u>not</u> a problem in Kenya	J7d
				Government is doing a good job ensuring that electricity is provided in Kenya *Note: Binary variable indicating "agree" or "strongly agree"*	J7g
9.6	Gender equality index	Resp.	Index	It is acceptable for a woman to be a bus driver	J6a
				Important decisions of the family should <u>not only</u> be made by the man of the family	J6b
				If the wife is working outside the home, the husband should help her with household chores	J6c
				Women should have more opportunities to become political leaders *Note: Binary variable indicating "agree" or "strongly agree"*	J6d
9.7	Ethnic identity index	Resp.	Index	Ethnic identity is "important" or "very important" in respondent's life	J4e

				Indicator for belongs first to ethnic group (over other dimensions of identity)	J4f
9.8	Religiosity index	Resp.	Index	Religious identity is "important" or "very important" in respondent's life Indicator for belongs first to religious group (over other dimensions of identity) Attends church/mosque regularly Attended church/mosque last week	J4d J4f J4a J4b
9.9	Social trust index	Resp.	Index	Trusts people, in general Trusts members of their own ethnic group Trusts members of other ethnic groups Trusts members of their own religion Trust members of other religions *Note: Indicator for "can be trusted" or "can be somewhat trusted"*	J5a J5b J5c J5d J5e

3.12 Family #10 – Community outcomes

There are a number of community-level outcomes that are of interest in this study. For example, Bernard and Torero (2015) find that take-up of electricity may be higher in communities where electricity is more prevalent. Therefore, a key outcome of interest in our study is whether the subsidy treatments impacted the proportion of secondary sample households choosing to connect to electricity. In addition, it is possible that electricity can lead to actual or perceived within-village inequality, in income, educational outcomes, and consumption. In order to estimate the impacts of electrification on within-community inequality, we will take advantage of our random sample of households and calculate Gini coefficients, capturing within-community dispersion, using the productivity (Family 4), wealth (Family 5), education (Family 8), and consumption (Family 6) outcomes in our data.[16]

Table 11. Community primary outcomes

ID	Outcome	Unit	Type	Details	Question
10.1	Community electrification rate	Com.	Proportion	Estimated community electrification rate	See Section 2.3
10.2	Community electricity reliability index	Com.	Index	Proportion of connected households reporting power blackouts in past 7 days Proportion of connected households reporting regular blackouts	F10c, F10d F10e
10.4	Value of assets inequality	Com.	Index	Gini coefficient capturing within-community dispersion in total asset value	See 5.6 above

[16] Note that we will weight observations according to their proportions (e.g. main sample, secondary sample, etc.) households in the baseline community census data.

10.5	Education inequality	Com.	Index	Gini coefficient capturing within-community dispersion in student test score results	T1, T2
10.6	Consumption inequality	Com.	Index	Gini coefficient capturing within-community dispersion in total consumption of 23 consumption goods	M5, M7, M8
10.7	Perceived income inequality	Com.	Proportion	Proportion of respondents agreeing with statement that economic inequality is a problem in this village	J7e

References

Anderson, Michael L. 2008. "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(484): 1481-1495.

Barron, Manuel, and Maximo Torero. 2015. "Household Electrification and Indoor Air Pollution."

Bernard, Tanguy and Maximo Torero. 2015. "Social Interaction Effects and Connection to Electricity: Experimental Evidence from Rural Ethiopia." *Economic Development and Cultural Exchange* 63(3): 459-484.

Bruhn, Miriam and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics* 1(4): 200-232.

Burlig, Fiona and Louis Preonas. 2016. "Out of the Darkness and Into the Light? Development Effects of Electrification in India."

Chakravorty, Ujjayant, Kyle Emerick, and Majah-Leah Ravago. 2016. "Lighting Up the Last Mile: The Benefits and Costs of Extending Electricity to the Rural Poor."

Dinkelman, Taryn. 2011. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." *American Economic Review* 101(7): 3078–3108.

Furukawa, Chishio. 2014. "Do Solar Lamps Help Children Study? Contrary Evidence from a Pilot Study in Uganda." *Journal of Development Studies* 50(2): 319-341.

Grimm, Michael, Anicet Munyehirwe, Jorg Peters, and Maximiliane Sievert. 2015. "A First Step Up the Energy Ladder? Low Cost Solar Kits and Household's Welfare in Rural Rwanda." *Ruhr Economic Paper* 554.

Grogan, Louise, and Asha Sadanand. 2012. "Rural Electrification and Employment in Poor Countries: Evidence from Nicaragua." *World Development* 43: 252-265.

Hassan, Fadi, and Paolo Lucchino. 2016. "Powering Education." *CEP Discussion Paper No. 1438*.

Khandker, Shahidur, Hussain Samad, Rubaba Ali, and Douglas Barnes. 2012. “Who Benefits Most from Rural Electrification? Evidence from India.” *World Bank Policy Research Working Paper* 6095.

Kowalski, Amanda E. 2016. “Doing More When You’re Running LATE: Applying Marginal Treatment Effect Methods to Experiments.”

Ligon, Ethan. 2015. “Estimating Household Neediness from Disaggregated Expenditures.”

Lipscomb, Molly, Mobarak, Ahmed Mushfiq, and Tania Barham. 2013. “Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil.” *American Economic Journal: Applied Economics* 5(2): 200-231.

Lee, Kenneth, Eric Brewer, Carson Christiano, Francis Meyo, Edward Miguel, Matthew Podolsky, Javier Rosa, and Catherine Wolfram. 2016. “Electrification for “Under Grid” Households in Rural Kenya.” *Development Engineering* 1: 26-35.

Lee, Kenneth, Edward Miguel, and Catherine Wolfram. 2016a. “Appliance Ownership and Aspirations among Electric Grid and Home Solar Households in Rural Kenya.” *American Economic Review: Papers & Proceedings* 106(5): 89-94.

Lee, Kenneth, Edward Miguel, and Catherine Wolfram. 2016b. “Experimental Evidence on the Demand for and Costs of Rural Electrification.” *NBER Working Paper* 22292.

McKenzie, David. 2012. “Beyond baseline and follow-up: The case for more T in experiments.” *Journal of Development Economics* 99(2): 210-221.

Pre-analysis plan C

"The Economic and Social Impacts of Electrification: Evidence from Kenya"[1]

AEA RCT Title: "Evaluation of Mass Electricity Connections in Kenya"

RCT ID: AEARCTR-0000350

Principal Investigators: Kenneth Lee (Energy Policy Institute at the University of Chicago), Edward Miguel (University of California, Berkeley), and Catherine Wolfram (University of California, Berkeley)

Date: 27 March 2018

Summary: This document outlines the plan for analyzing a dataset consisting of information on the living standards of roughly 2,500 households in Western Kenya, including nearly 500 households that previously benefited from a randomized household electrification program. The goal of this study is to estimate the economic and social impacts of household electricity connections. This document lays out the main regression specifications and outcome variable definitions that we intend to follow. However, we anticipate that we will carry out additional analyses beyond those included in this document. Therefore, this document is not meant to be comprehensive or to preclude additional analyses.

Appendices:

- A. Living Standards Kenya (LSK) Survey – Baseline (2014)
- B. Living Standards Kenya (LSK) Survey – Follow-up Round 1 (2016)
- C. Living Standards Kenya (LSK) Survey – Follow-up Round 2 (2017)
- D. Note on data management/access and pre-analysis plan (uploaded in October 2017)

[1] This research is supported by the Berkeley Energy and Climate Institute, the Blum Center for Developing Economies, the Center for Effective Global Action, the Development Impact Lab (USAID Cooperative Agreements AID-OAA-A-13-00002 and AIDOAA-A-12-00011, part of the USAID Higher Education Solutions Network), the International Growth Centre, the U.C. Center for Energy and Environmental Economics, the Weiss Family Program Fund for Research in Development Economics, the World Bank DIME i2i Fund, and an anonymous donor. Corresponding author: Edward Miguel (emiguel@berkeley.edu).

1. Introduction

1.1 Summary

Universal access to modern energy has become a top priority for policymakers, nongovernmental organizations, and international donors across Sub-Saharan Africa. In Kenya, nearly US$600 million has been invested in extending the grid to rural areas since 2008. While there is now widespread grid coverage, the national household electrification rate remains relatively low. Kenya is currently pursuing a strategy of last-mile connections for "under grid" households in order to reach universal access to electricity by 2020. Given the high cost of subsidizing mass connections, however, there is a need for better understanding of the impacts of rural electrification. *In this study, we will provide experimental evidence on the impacts of household electrification across a range of economic and social outcomes in Western Kenya.*

Between 2013 and 2015, we implemented a field experiment in which electricity connection vouchers (worth varying amounts) were randomly assigned to clusters of rural households in Western Kenya. Households accepting these vouchers were then connected to the national grid, in cooperation with Kenya's Rural Electrification Authority (REA) and Kenya Power, the main electricity distribution company. As a result of this experiment, it is possible to perform a randomized evaluation of the impact of household grid connections. The study focuses on household survey data from a baseline survey and two follow-up surveys of roughly 2,294 households that were observed to be unconnected at baseline.

1.2 Experimental design and steps

In this section, we describe the experimental design. For further details, see Lee et al. (2016) at http://dx.doi.org/10.1016/j.deveng.2015.12.001, Lee, Miguel, and Wolfram (2016) at http://dx.doi.org/10.1257/aer.p20161097, and Lee, Miguel, and Wolfram (2018) at http://www.nber.org/papers/w22292. (This third paper has been submitted for journal publication, and the final published version will differ somewhat from the earlier NBER working paper version that we link to here.)

Step 1: In July 2013, we collaborated with REA to identify a list of 150 rural "transformer communities" that would form a representative sample of communities recently connected to the electrical grid in Busia and Siaya, two counties in Western Kenya. Each community is defined as all of the structures that were located within 600 meters of a central transformer.

Step 2: Between September 2013 and December 2013, we visited each community and geo-tagged over 13,000 structures, capturing the universe of un-electrified households that could potentially be connected to the national grid.

Step 3: Using these data as a sampling frame, we randomly sampled 2,504 households, consisting of 2,294 households that were unconnected at baseline and 210 households that were connected to the grid at baseline. The regressions described in Sections 2.2 to 2.4 focus on the group of 2,294 households. We use data from the sample of 210 connected households mainly for descriptive purposes, for example, to compare characteristics of households that had already connected without our subsidy to households that later connected with a subsidy. Between February and August 2014, we administered a detailed survey of each household, capturing baseline measures of living standards. See "Living Standards Kenya (LSK) Survey – Baseline (2014)" in Appendix A.

Step 4: In April 2014, we randomly assigned the 150 communities into four groups: (1) "High-subsidy" (or 100% discount) arm with 25 communities, resulting in an effective price of $0; (2) "Medium-subsidy" (or 57% discount) arm with 25 communities, resulting in an effective price of $171; (3) "Low-subsidy" (or 29% discount) arm with 25 communities, resulting in an effective price of $284, and (4) "No subsidy" or control group (effective status quo price of $398 plus wiring) with 75 communities.

Step 5: After distributing the electricity connection subsidies, we facilitated the construction of grid infrastructure to connect the 478 unconnected households that accepted the randomized offer. The first household was metered in September 2014, the average connection time was seven months, and the final household was metered over a year later, in October 2015.

Step 6: Between May and September 2016, we administered a follow-up survey ("Follow-up Round 1") and successfully surveyed 2,217 study households, or 96.6 percent of baseline sample. We surveyed an additional 1,345 households—or six to eleven households per community—as part of a "spillover" or "secondary sample," randomly sampling households that were observed to be unconnected at the time of the census but were not enrolled during the baseline survey.

Data from this secondary sample is used to study within-village externality impacts to local households. We also collected follow-up survey data from 208 of the 210 households that had already been connected at the time of the baseline census. As part of the Follow-up Round 1 survey, we administered short English and Math tests to all 12 to 15 year olds in the sample households, or 2,317 children in total. See "Living Standards Kenya (LSK) Survey – Follow-up Round 1 (2016)" in Appendix B.

Step 7: In October 2017, we launched a second follow-up survey ("Follow-up Round 2") to capture various measures of living standards and other outcomes targeting the 2,504 households sampled during the baseline round. See "Living Standards Kenya (LSK) Survey – Follow-up Round 2 (2016)" in Appendix C. The Follow-up Round 2 survey was carried out between October and December 2017.

1.3 Main and secondary samples

To summarize, our study will focus on two sets of households. The first set of households—which we refer to as "main sample" households—consists of the 2,294 households that were unconnected to electricity at the time of the baseline survey. The second set of households—which we refer to as "spillover" or "secondary sample" households—consists of the 1,345 households that were enrolled into the data collection during the 2016 follow-up survey round. The analyses described in this pre-analysis plan focus only on main sample households, and the Follow-up Round 2 survey data.

1.4 Analysis and data examined to date

At the time of registering this pre-analysis plan, we had completed data collection for the Follow-up Round 2 surveys. Note that we did not examine the data or perform any data analysis before registering this plan. As described in the document titled, "Note on data management/access and pre-analysis plan," which was uploaded to the AEA RCT Registry in October 2017 and is included in Appendix D of this document, the authors of this pre-analysis plan were provided with access to de-identified survey data for roughly 250 surveys, at the very beginning of the survey round. These data were stripped of any indicators that could expose the treatment status of households, and were provided in order to (1) allow the authors to identify and correct any coding errors in the survey instrument, (2) make improvements to the choice sets for multiple-choice questions, (3) identify and amend questions that were taking excessive time to administer, (4) address any other technical issues with the survey instrument (for instance, with the SurveyCTO data entry software coding), and (5) make any final additions to the survey instrument to address minor questions that came up in the field. 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 plan to analyze these data. However, we anticipate carrying out additional tests. In other words, this plan is not meant to be an exhaustive set of all analyses, but rather a core set of important initial estimates that will hopefully inspire further analyses.

2. Analysis

2.1 General notes

Randomly lowering the price of an electricity connection at the community-level by 29, 57, and 100 percent, resulted in increases in take-up of 6%, 22%, and 95%, over the baseline, respectively.[2] Take up in the low and medium subsidy treatment arms was relatively low. In our analysis, we will estimate both treatment-on-treated (TOT) and intention-to-treat (ITT) impacts of electrification. ITT estimates will be obtained from specifications in which various outcomes of interest are regressed on a set of binary variables indicating the treatment status of the community. TOT estimates will be obtained from two-stage least squares specifications in which the household's electrification status is instrumented with the set of treatment indicators.

Throughout this document, we refer to our subject population as "households." In our setting, residential structures are typically located in compounds that can sometimes consist of multiple households. Our subject population consists of households that were considered to be the "main household" in the residential compound at the time of the baseline survey. To construct our sample, we randomly sampled compounds from each transformer community and enrolled

[2] See Lee, Miguel, and Wolfram (2018) for details.

the primary household in the compound. All other households in each compound are referred to as "minor households."

In the majority of our analysis, we will focus on the family of the respondent that was interviewed at baseline, regardless of whether the family is still living in the same location at the time of the follow-up survey. In practice, residential out-migration rates were low during the study period. For certain outcomes, however, we will focus on the family (if any) that is currently living at the physical location where the baseline survey took place. This will allow us to examine an additional set of questions including, for example, whether locations that were electrified are more likely to remain inhabited, compared to locations that were not electrified.

In Sections 2.2 to 2.4, we describe three analytical approaches. First, focusing on Round 2 cross-sectional data, we will estimate the impacts of grid electrification. Second, pooling together Round 1 and Round 2 data, we will perform largely the same analyses; using both rounds of follow-up data leads to improved statistical power. Finally, wherever we have equivalent outcome measures collected across the Baseline, Follow-up Round 1, and Follow-up Round 2 datasets, we will estimate the panel regressions described below.

2.2 Cross-sectional results

Using only the most recent Follow-up Round 2 survey data, we will test the hypothesis that households connected to the electricity grid enjoy higher levels of living standards. Specifically, we will estimate ITT results using the following equation:

$$y_{ic} = \beta_0 + \beta_1 T_{Lc} + \beta_2 T_{Mc} + \beta_3 T_{Hc} + X_c'\Lambda + Z_{ic}'\Gamma + \epsilon_{ic} \tag{1}$$

where y_{ic} represents the outcome of interest for main sample household i in community c, and T_{Lc}, T_{Mc}, and T_{Hc} are binary variables indicating whether community c was randomly assigned into the low-value, medium-value, and high-value subsidy treatment arms, respectively. Following Bruhn and McKenzie (2009), we include a vector of community-level characteristics, X_c, containing the variables used for stratification during randomization. In addition, we include Z_{ic}, a vector of household-level characteristics. Further details on the components of the covariate vectors are presented in Section 2.8. The variables in Z_{ic} will sometimes be used in analyses of treatment effect heterogeneity, which is discussed in further detail in Section 2.9. In Section 2.12, we discuss the possibility of ANCOVA specifications for certain outcome variables. In all cases, standard errors will be clustered at the community level.

The issue of limited statistical power may be more severe in ITT specifications due to the relatively low take-up rates in the low and medium subsidy treatment groups. To address this issue, we will focus attention on the coefficient on the high subsidy treatment indicator. This test will not only shed light on the impacts of near universal electrification (compared to the control group, were very few households become connected), but also is likely to have greater statistical power.

We will also estimate TOT results using the following equations:

$$E_{ic} = \delta_0 + \delta_1 T_{Lc} + \delta_2 T_{Mc} + \delta_3 T_{Hc} + X_c'\Lambda_1 + Z_{ic}'\Gamma_1 + \eta_{ic} \tag{2}$$

$$y_{ic} = \beta_0 + \beta_1 E_{ic} + X_c'\Lambda_2 + Z_{ic}'\Gamma_2 + \epsilon_{ic} \tag{3}$$

where the first-stage equation 2 estimates the effects of the treatment indicators on household electrification status, E_{ic}, and the second-stage equation 3 estimates the effect of household electrification status on the various outcomes of interest. As in equation 1, errors will be clustered at the community level.

Lee, Miguel, and Wolfram (2018) document systematic differences in the baseline living conditions of households taking up the experimental offers in the low and medium subsidy groups, compared to the high subsidy group. Households that paid more for an electricity connection (i.e., lower subsidy arm households) were wealthier and more educated on average than those who paid nothing (i.e., high subsidy arm households). This suggests that the average treatment effect may vary across treatment arms. For example, electrification may be more impactful for the relatively wealthier households that are able to invest in complementary assets such as electrical appliances. In order to examine these types of heterogeneous treatment effects, we will explore the methods described in Kowalski (2016) to first recover bounds on average treatment effects for "always taker" and "never taker" households, and then decompose group average treated outcomes into selection and treatment effects. However, due to relatively low take-up rates in the low and medium subsidy groups, some of these analyses may be statistically underpowered.

In our analyses of Follow-up Round 1 data, some of which are described in Lee, Miguel, and Wolfram (2018), we find little to no evidence of spillovers accruing to local households. We can therefore interpret the TOT results largely without concern about violations of the stable unit treatment value assumption (SUTVA).

2.3 Pooled results

In order to improve statistical power, we will pool together Follow-up Round 1 and Round 2 data, and estimate ITT results using the following equation:

$$\mathrm{y}_{\mathrm{ict}} = \beta_0 + \beta_1 \mathrm{T}_{\mathrm{Lc}} + \beta_2 \mathrm{T}_{\mathrm{Mc}} + \beta_3 \mathrm{T}_{\mathrm{Hc}} + \mathrm{X}_{\mathrm{c}}' \Lambda + \mathrm{Z}_{\mathrm{ict}}' \Gamma + \mu_{\mathrm{R2}} + \epsilon_{\mathrm{ict}} \tag{4}$$

In addition, we will estimate TOT results using the following equations:

$$\mathrm{E}_{\mathrm{ict}} = \delta_0 + \delta_1 \mathrm{T}_{\mathrm{Lc}} + \delta_2 \mathrm{T}_{\mathrm{Mc}} + \delta_3 \mathrm{T}_{\mathrm{Hc}} + \mathrm{X}_{\mathrm{c}}' \Lambda_1 + \mathrm{Z}_{\mathrm{ict}}' \Gamma_1 + \mu_{\mathrm{R2}} + \eta_{\mathrm{ict}} \tag{5}$$

$$\mathrm{y}_{\mathrm{ict}} = \beta_0 + \beta_1 \mathrm{E}_{\mathrm{ict}} + \mathrm{X}_{\mathrm{c}}' \Lambda_2 + \mathrm{Z}_{\mathrm{ict}}' \Gamma_2 + \mu_{\mathrm{R2}} + \epsilon_{\mathrm{ict}} \tag{6}$$

In the above equations, μ_{R2} denotes Follow-up Round 2 data. Equation 5 will effectively estimate the average take-up effects over both follow-up rounds, and Equation 4 and Equation 6 will similarly estimate the average electrification treatment effects on the outcomes of interest over both rounds, for those outcome measures that were collected in both Follow-up Round 1 and Round 2.

2.4 Panel results

Finally, we will combine Follow-up Round 1 and Follow-up Round 2 data into a panel and focus on the set of outcome measures that are available across all rounds. Specifically, we will estimate ITT results using the following equation:

$$\begin{aligned}\mathrm{y}_{\mathrm{ict}} = {} & \beta_{01} D_t^{R1} + \beta_{02} D_t^{R2} + \beta_{11} \mathrm{T}_{\mathrm{Lct}} D_t^{R1} + \beta_{21} \mathrm{T}_{\mathrm{Mct}} D_t^{R1} + \beta_{31} \mathrm{T}_{\mathrm{Hct}} D_t^{R1} + \\ & \beta_{12} \mathrm{T}_{\mathrm{Lct}} D_t^{R2} + \beta_{22} \mathrm{T}_{\mathrm{Mct}} D_t^{R2} + \beta_{32} \mathrm{T}_{\mathrm{Hct}} D_t^{R2} + D_t^{R1} \mathrm{X}_{\mathrm{c}}' \Lambda_1 + \\ & D_t^{R1} \mathrm{Z}_{\mathrm{ict}}' \Gamma_1 + D_t^{R2} \mathrm{X}_{\mathrm{c}}' \Lambda_2 + D_t^{R2} \mathrm{Z}_{\mathrm{ict}}' \Gamma_2 + \epsilon_{\mathrm{ict}}\end{aligned} \tag{7}$$

In addition, we will estimate TOT results using the following equations:

$$\begin{aligned}\mathrm{E}_{\mathrm{ict}} = {} & \delta_{01} D_t^{R1} + \delta_{02} D_t^{R2} + \delta_{11} T_{Lct} D_t^{R1} + \delta_{21} T_{Mct} D_t^{R1} + \delta_{31} T_{Hct} D_t^{R1} + \\ & \delta_{12} T_{Lct} D_t^{R2} + \delta_{22} T_{Mct} D_t^{R2} + \delta_{32} T_{Hct} D_t^{R2} + D_t^{R1} \mathrm{X}_{\mathrm{c}}' \Lambda_{1,1} + \\ & D_t^{R1} \mathrm{Z}_{\mathrm{ict}}' \Gamma_{1,1} + D_t^{R2} \mathrm{X}_{\mathrm{c}}' \Lambda_{2,1} + D_t^{R2} \mathrm{Z}_{\mathrm{ict}}' \Gamma_{2,1} + \eta_{\mathrm{ict}}\end{aligned} \tag{8}$$

$$y_{ict} = \beta_{01} D_t^{R1} + \beta_{02} D_t^{R2} + \beta_1 E_{ict} D_t^{R1} + \beta_2 E_{ict} D_t^{R2} + D_t^{R1} X_c' \Lambda_1 + D_t^{R1} Z_{ict}' \Gamma_1 + D_t^{R2} X_c' \Lambda_2 + D_t^{R2} Z_{ict}' \Gamma_2 + \epsilon_{ict} \quad (9)$$

In the above equations, D_t^{R1} is an indicator variable that equals one for Follow-up Round 1 observations and zero otherwise. Similarly, D_t^{R2} equals one for Follow-up Round 2 observations and zero otherwise. In the ITT specification, we will focus attention on the two coefficients on the high subsidy treatment indicators (β_{31} and β_{32}). We will test the null hypothesis that both coefficients are equal to zero to assess whether electrification had impacts in any period. We will also test the null hypothesis that the two coefficients are equal to each other to assess whether the effects of electrification differed over time. Similarly, in the TOT specification (equation 9), we will focus on β_1 and β_2, and will both test the null hypothesis that both coefficients are equal to zero to assess whether electrification had impacts in any period, and we will also test the hypothesis that the two coefficients are equal to each other.

2.5 Community-level outcomes

For community-level outcomes (which are specified in Section 3.12), we will estimate equations that are similar in form to those specified in Sections 2.2 to 2.4. Since the unit of observation is the community, we will exclude household-level covariates. When using the Follow-up Round 2 data on community outcomes in the TOT specification, we will replace the E_{ic} term in equations 2 and 3 with R_c, the estimated local transformer community total electrification rate.

Note that for each transformer community, we have data on the universe of households (as well as their grid connection status) at the time of our baseline census. In addition, we have Round 1 follow-up household survey data for the main and secondary sample households and Round 2 follow-up survey data for the main sample households. Since we do not have updated census data for each transformer community, we will need to estimate the current electrification rate. For each of the treatment arms, we will calculate the average take-up rate at the time of the Round 1 follow-up for the portion of secondary sample households that were observed to be unconnected at the time of the baseline census. We will estimate R_c by combining actual follow-up take-up (connection) data among the main sample households surveyed in Round 2 with estimated connection data for the non-surveyed households in the relevant treatment group.

Specifically, for each treatment arm, we will assume that all of the remaining, non-surveyed households connected to the grid at the treatment arm-level average take-up rate for non-surveyed households (i.e., the households that did not receive connection subsidies even in the subsidy treatment arms). We will estimate the average take-up rate using Round 1 follow-up data for the spillover sample. Due to concerns that the TOT specification may be statistically underpowered, for the community-level outcomes we will focus primarily on the ITT results. See Section 3.12 for additional details on how we plan to construct community-level outcome variables.

2.6 Educational impacts

Another objective of this study is to understand the extent to which household electrification impacts the educational outcomes of schoolchildren. As part of the Follow-up Round 2 survey, we collected student scores on the Kenya Certificate of Primary Education (KCPE) exam, a standardized national exam administered at the end of primary school and required for admission to secondary school. Using these data, we will estimate regressions that are similar in form to those specified in Sections 2.2 to 2.4 but will focus on individual children as the unit of observation. In these regressions, the covariate vector $\mathrm{Z_{ic}}$ will be complemented with the covariate vector $\mathrm{C_{jic}}$, which includes additional information on child j in household i in community c (e.g., child demographic characteristics). The outcomes of interest in these specifications will therefore be denoted with the subscript *jic*. The covariate vector $\mathrm{C_{jic}}$ is described in more detail in Section 2.8.

2.7 Stated willingness to pay (WTP) for electricity

In the Round 2 follow-up survey, we first ask respondents whether they would be hypothetically willing to connect to the national grid at a randomly selected price (i.e., in a *time unlimited* offer) (f6g in the Round 2 survey). The randomly selected price, p, was drawn from the following set of prices (in Kenyan shillings):

$$\{0, 5000, 10000, 15000, 20000, 25000, 35000, 75000\}$$

This question was followed by an additional hypothetical question (f3h) asking the respondent whether they would accept an offer at this price if they were given only six weeks to complete the payment (i.e., *time limited* offer). Finally, respondents were asked whether they would be willing to pay a monthly amount over a period of three years, where the cumulative total is equal to the randomly selected price (f3i) (i.e., *financed offer*, with terms similar to those offered under the current Kenya LMCP). Respondents from connected households were asked a similar set of questions with somewhat different wording to reflect the fact that they are already connected (see f4d, f4e, and f4f).

We are interested in understanding how stated WTP responds to price levels. Specifically, we will estimate the following equation:

$$\begin{aligned} h_{ic} = \ & \alpha_0 + \alpha_1 T_{Lc} + \alpha_2 T_{Mc} + \alpha_3 T_{Hc} + \textstyle\sum_p \beta_p W_{pic} + \sum_p \gamma_{1p}(W_{pic} \times T_{Lc}) + \\ & + \textstyle\sum_p \gamma_{2p}(W_{pic} \times T_{Mc}) + \sum_p \gamma_{3p}(W_{pic} \times T_{Hc}) + X_c'\Lambda + Z_{ic}\Gamma + \epsilon_{ic} \end{aligned} \tag{10}$$

where h_{ic} is a binary variable indicating the stated (i.e., hypothetical) take-up decision for household i, W_{pic} is a binary variable indicating whether household i in community c received the hypothetical price p, and Z_{ic} is the household covariate vector. We are especially interested both in the direct effects of the treatment indicators, as well as the coefficients on the full set of interactions between the treatment indicators and the W_{pic} terms. These interactions will shed light on how stated WTP may differ for households that were recently connected to the grid (e.g., using the main sample data), and thus may have direct experience with household electrification, or for unconnected households that recently observed neighboring households become connected to the grid (e.g., using the secondary sample data), although this second analysis is only possible using the Follow-up Round 1 data. Standard errors will be clustered at the community level.[3] We will also test for heterogeneous effects, which are generally described in Section 2.9.

As in Lee, Miguel, and Wolfram (2018), we will plot the stated WTP results graphically. For example, we may plot and compare demand curves for (1) time unlimited, time limited, and financed offers, (2) control households at baseline versus at each follow-up round, and (3) main sample households in the various subsidy arms versus the control group at baseline and at each follow-up round, as well as other comparisons.

[3] Based on the results of Lee, Miguel, and Wolfram (2018), we do not expect the relationship between take-up and price to be linear. However, we may still test for linearity, and if we cannot reject linearity in an F-test, we will also estimate an equation in which y_{ic} is regressed on p_{ic}, controlling for the treatment indicators and other covariates.

2.8 Covariate vectors X_c, Z_{ic}, and C_{jic}

In this section, we describe each of the sets of covariates that we plan to utilize in the analysis.

The vector $\mathrm{X_c}$ will primarily include the stratification variables that were used during randomization. These include:

- County: Binary variable indicating whether community c is in Busia county or Siaya county.
- Market status: Binary variable indicating whether the total number of businesses in community c is strictly greater than the community-level mean across the entire sample at baseline. We use this definition to define which communities could reasonably be classified as “markets” relative to others.
- Transformer funding year: Binary variable indicating whether the electricity transformer in community c was funded “early” (i.e. in either 2008-09 or 2009-10).
- Electrification rate: Residential electrification rate in community c at the time of census (roughly 2013).
- Community population: Estimated number of people living in community c at the time of census (roughly 2013).

The vector $\mathrm{Z_{ic}}$ will include the set of household-level variables listed below, taking advantage of the baseline survey data.

- Gender of respondent: Binary variable indicating whether the baseline respondent was female.[4]
- Age: Age of the baseline respondent in 2014.
- Education of respondent at baseline: Binary variable indicating whether the household respondent at baseline has completed secondary school.[5]
- Bank account at baseline: Binary variable indicating whether the household respondent at baseline had a bank account.

[4] As a robustness check, we will also present a specification in which we control for the demographic characteristics of the respondent (e.g., gender and age) in the follow-up survey.

[5] The respondent during the baseline survey is not necessarily the same person as the respondent during the follow-up survey.

- Housing quality index at baseline: Index composed of whether the household had high-quality floors, roof, and walls at baseline.
- Asset value at baseline: Estimated value based on inventory of livestock, electrical appliances, and non-livestock assets at baseline, at current observed local prices.
- Energy spending at baseline: Estimated monthly expenditures on all energy sources at baseline.

The vector C_{jic} will include a set of individual-level characteristics that are relevant for the regression specifications estimating the impacts of electrification on educational performance.

- Gender of student: Binary variable indicating whether the student is female.
- Age: Age of student in 2017.
- Siblings: Number of children under 18 in the household.
- Grade attained at baseline: Grade attained by the end of the 2013 academic year.[6]

2.9 Heterogeneous effects

In additional analyses, we will estimate heterogeneous treatment effects along a number of major dimensions, captured in the vectors X_c, Z_{ic}, and C_{jic}, by adding interaction terms between each treatment indicator and these variables. For instance, in order to assess how treatment impacts may vary for households at different wealth levels, we will estimate specifications in which the treatment indicators are interacted with the housing quality index at baseline.

Furthermore, there are additional variables that are not included in the covariate vectors above but are of potential interest. One variable of interest is the frequency of transformer outages in the community—that is, the proportion of months (between September 2014 and June 2017) that the transformer was not working. Intuitively, it appears likely that any electrification impacts would be muted in communities without a functioning transformer.

We are uncertain whether our study design will have sufficient statistical power to generate precise estimates on many of these interaction terms and hence such analyses should be

[6] We will infer this data by comparing the baseline and follow-up surveys for main sample households. It is possible that this data will be missing for some individuals. If there are relatively few such instances, we will include an additional binary variable as a covariate indicating that the data are missing. Alternatively, we will likely drop this covariate altogether if this data on grade attainment in 2013 is missing for over 30% of individuals in the data.

considered suggestive rather than definitive. The patterns that emerge will also likely stimulate further exploratory analysis using the dataset.

2.10 Construction of indices

When constructing indices, we will normalize each component variable to have mean zero and unit variance, and thereafter we will construct the index by summing each component variable and then re-normalizing (the mean effects approach). Note that we will exclude any variables with zero (or very close to zero) variance since these do not contribute any information to the analysis. Furthermore, if a pre-specified variable is missing for more than 30% of possible observations collected in the follow-up surveys, we will drop it from inclusion in the index. We cannot anticipate why a particular variable will be missing so frequently, and believe such cases will be rare, but in such events where it warrants exclusion, we shall also explore these reasons in the analysis. Finally, in the appendix we will also report results for all individual outcomes used to create indices.

2.11 Multiple Testing Adjustment

In Section 3, we describe how the major outcomes of interest are categorized into eleven (11) broad "families". For the main coefficient estimates of interest (for instance, β_1, β_2, and β_3 in equation 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. For instance, researchers interested in the effect of household electrification on non-agricultural earnings should focus directly on this p-value.

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 null hypothesis.[7] 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 outcomes within the primary outcomes group (Table 1), as well as FDR-adjusted q-values for each outcome within each of the eleven outcome families (Tables 3 through 13). 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.

2.12 Additional analyses

For a subset of outcomes in the main sample regressions, we will have comparable measures in the baseline survey as well as in both Follow-up Rounds (e.g., we have such data for household size, home solar system usage, energy consumption, etc.). In these cases, we are also able to estimate ANCOVA regression specifications in which the baseline value of the outcome of interest is included as an additional covariate, as the resulting estimates may have greater statistical power (McKenzie 2012). However, note that we lack equivalent baseline measures for most outcome variables described below (in Section 3). As a result, the ANCOVA estimates will be presented mostly as a supplement to the analyses already described, and our main focus will be on the results of the specifications described in Sections 2.2 and 2.4 above.

3. Major outcomes of interest

3.1 Overview

In this section, we specify 85 major economic and social outcomes of interest. These outcomes have been selected based on the judgment of the research team and are arranged into eleven broad families: (1) energy consumption, (2) household structure, (3) time use, (4) productivity, (5) wealth, (6) consumption, (7) health and wellbeing, (8) education, (9) social and political attitudes, (10) community outcomes, and (11) safety and crime outcomes. Based on this list, we also identify a group of eleven "primary" outcomes, as well as two "grouped" outcomes, drawn from a number of different outcome families. The estimated impacts on these primary

[7] 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.

outcomes will serve as an overall summary of the impacts of household electrification in our setting.

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

Table 1 summarizes the primary outcomes that will serve as an overall summary of living standards and life outcomes in our setting. For certain primary outcomes, we are able to use the existing literature to guide our expectations on the impacts of electrification in our setting. For example, in South Africa, Dinkelman (2011) finds that female employment rises by 9 to 9.5 percentage points and women work roughly 8.9 hours more per week. In Brazil, Lipscomb, Mobarak, and Barham (2013) find that the probability of employment increases by 17 to 18 percentage points, over the long run, in counties that are electrified. Taken together, we should expect to find substantial increases in the probability of employment (P.3a) and labor hours (P.4), particularly for women. Furthermore, in the Philippines, Chakravorty, Emerick, and Ravago (2016) estimate that village-level electrification leads to an increase in household expenditures by 38 percent, suggesting that there will be large gains in household consumption (P.6). In terms of test scores (P.9a, P.9b), Hassan and Lucchino (2016) examine the impacts of randomly distributing solar lanterns to 7th grade pupils in Kenya and find math grades to increase by 0.88 standard deviations for treatment pupils. In our analysis of each primary outcome, we will test the null hypothesis of no effect, and (wherever possible) the hypothesis that the treatment effect is the same as that found in the existing literature. Finally, we will compare the estimated impacts in our study to other outcomes in the broader development economics literature in order to assess the cost effectiveness of rural electrification as a development policy.

Table 2 presents two groupings of the primary outcomes in Table 1. The primary economic outcomes (P3-P6) are combined into an economic mean effect index. The primary non-economic outcomes (P7-P11) are combined into a non-economic mean effect index. We will use these indices to test whether electrification had an overall economic effect or non-economic effect on households.

3.3 Family #1 – Energy consumption major outcomes

At the most basic level, electricity connections should impact the ways in which households consume energy. Family 1 outcomes are presented in Table 3.

3.4 Family #2 – Household structure major outcomes

If there are changes in the patterns of energy consumption, there may also be changes in the structure of the household. For example, access to electricity may impact household structure by influencing incentives to migrate by making living in the household more attractive. Family 2 outcomes are presented in Table 4.

3.5 Family #3 – Time use major outcomes

Household electrification may operate as a labor-saving technology shock to home production, releasing female time from home to market work (Dinkelman 2011; Grogan and Sadanand 2012). Family 3 outcomes are presented in Table 5. Note that we did not collect Family 3 outcomes in Follow-up Round 2.

3.6 Family #4 – Productivity major outcomes

If electrification changes people's time use, and, for example, allows for more hours of work outside the home, there may be positive impacts on various measures of productivity and wealth.[8] The evidence on the impacts of electrification on productivity have been somewhat mixed. Dinkelman (2011), for example, finds evidence of increased female labor force participation in South Africa. Chakravorty, Emerick, and Ravago (2016) find large impacts of electrification on household income and expenditures in the Philippines, but attribute these

[8] Grimm et al. (2015), for instance, present a theoretical model in which an increase in household electrification effectively reduces the price of energy faced by the household, which increases the productivity of domestic labor and the output of household production.

impacts to increases in agricultural income rather than increases in labor force participation. In contrast, Burlig and Preonas (2016) find little to no impacts of electrification on various employment outcomes in rural India. Family 4, shown in Table 6, includes various measures of household agricultural activities, employment, small businesses, and other outcomes. In Follow-up Round 2, it includes a measure of total household earnings, estimated by adding together net earnings for each household member from employment or self-employment as well as from agricultural activities.

3.7 Family #5 – Wealth major outcomes

In terms of wealth, Lipscomb, Mobarak, and Barham (2013) find evidence of higher average housing values as a result of electrification in Brazil. Family 5, shown in Table 7, includes a housing quality index and estimated values of different types of household assets, based on current market prices.

3.8 Family #6 – Consumption major outcomes

We are interested in estimating the impacts of electrification on various measures of household consumption. In our setting, we focus on 23 items, including staples, vegetables, meat, fruits, and other goods. These 23 items were identified using data from the Kenya Life Panel Survey (KLPS-3).[9] Based on the KLPS-3 data, the 23 items account for 26% of total household consumption and 52% of total food consumption. Family 6, shown in Table 8, summarizes the various consumption outcomes.

3.9 Family #7 – Health and wellbeing outcomes

Electricity has been found to improve respiratory health by reducing indoor air pollution (Barron and Torero 2017). Some people may also be happier when they have access to electricity due to impacts on various channels. Family 7, shown in Table 9, includes various measures of respondent health and wellbeing. Note that Family 7 data was collected only during the Follow-up Round 1 survey.

---

[9] The KLPS-3 project is located in the same study region as this project and is led by Edward Miguel and other researchers. In the full KLPS survey, respondents are asked in detail about their consumption of 153 items.

3.10 Family #8 – Education outcomes

It is possible that electrification may improve educational outcomes for students, if better lighting allows for more evening study time, for instance. The evidence, however, has been somewhat mixed to date. Randomized trials, including Furukawa (2014) and Hassan and Lucchino (2016), have focused on measuring the impacts of decentralized power solutions, such as solar lanterns, and have documented results ranging from negative impacts to positive impacts with substantial spillovers. Studies on the impacts of grid connections have been mostly non-experimental and have found positive impacts of electrification on school enrollment, study time, and school completion (see, e.g., Khandker et al. 2014). Family 8 includes a variety of educational outcomes, including test scores from English and Math tests that were administered to students in the sample villages by our project field staff.

The data include two different measures of test scores. The first test scores come from English and Math tests administered by our project field staff as part of Follow-up Round 1 data collection. The second test scores come from the KCPE exam, normally administered to students in their final year of primary school education. If we do not find evidence that electrification affected likelihood of taking the KCPE (outcome 8.9), then we intend to pool data for the average of math and English test scores (outcome 8.3) and KCPE scores (outcome 8.10) to be used as a single outcome variable. If electricity connections affected the likelihood of a connection, then interpretation of the average differences in Round 2 KCPE test scores will be more difficult to interpret, due to possible selection into the test score sample. Family 8 outcomes are listed in Table 10.

3.11 Family #9 – Social and political attitudes outcomes

Electrified households may consume more media content (via televisions, radios, and internet access), and as a result, could have greater knowledge of current affairs, or experience changes in social and political attitudes. Family 9 outcomes are listed in Table 11.

3.12 Family #10 – Community outcomes

There are a number of community-level outcomes that are of interest in this study. It is possible that electricity can affect actual or perceived within-village inequality, in income, educational outcomes, and consumption. In order to estimate the impacts of electrification on within-community inequality, we will take advantage of our random sample of households and calculate Gini coefficients, capturing within-community dispersion, using the productivity (Family 4), wealth (Family 5), education (Family 8), and consumption (Family 6) outcomes in our data. We will also match each transformer community to the nearest polling center from the August 2017 Kenyan national general elections. This will allow us to examine whether electrification affected local election vote share results.[10] Family 10 outcomes are listed in Table 12.

3.13 Family #11 – Safety and crime outcomes

Access to electricity may help reduce local crime. For example, electric lights inside or outside the home may improve safety by making attempts to engage in criminal behavior more visible. Family 11 outcomes are listed in Table 13.

References

Anderson, Michael L. 2008. "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(484): 1481-1495.

Barron, Manuel, and Maximo Torero. 2017. "Household Electrification and Indoor Air Pollution." *Journal of Environmental Economics and Management* 86: 81-92.

Bernard, Tanguy and Maximo Torero. 2015. "Social Interaction Effects and Connection to Electricity: Experimental Evidence from Rural Ethiopia." *Economic Development and Cultural Exchange* 63(3): 459-484.

Bruhn, Miriam and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics* 1(4): 200-232.

Burlig, Fiona and Louis Preonas. 2016. "Out of the Darkness and Into the Light? Development Effects of Electrification in India."

Chakravorty, Ujjayant, Kyle Emerick, and Majah-Leah Ravago. 2016. "Lighting Up the Last Mile: The Benefits and Costs of Extending Electricity to the Rural Poor."

Dinkelman, Taryn. 2011. "The Effects of Rural Electrification on Employment: New Evidence

[10] Note that for the analysis of voting outcomes, we plan to weight each community by the ratio of inhabitants living within 600 meters of the transformer to the total number of registered voters at the nearest polling center. In cases where only a small share of voters live near our sample transformer, the overall polling station data is less informative about the households we study.

from South Africa." *American Economic Review* 101(7): 3078–3108.

Furukawa, Chishio. 2014. "Do Solar Lamps Help Children Study? Contrary Evidence from a Pilot Study in Uganda." *Journal of Development Studies* 50(2): 319-341.

Grimm, Michael, Anicet Munyehirwe, Jorg Peters, and Maximiliane Sievert. 2015. "A First Step Up the Energy Ladder? Low Cost Solar Kits and Household's Welfare in Rural Rwanda." *Ruhr Economic Paper* 554.

Grogan, Louise, and Asha Sadanand. 2012. "Rural Electrification and Employment in Poor Countries: Evidence from Nicaragua." *World Development* 43: 252-265.

Hassan, Fadi, and Paolo Lucchino. 2016. "Powering Education." *CEP Discussion Paper No. 1438*.

Khandker, Shahidur, Hussain Samad, Rubaba Ali, and Douglas Barnes. 2014. "Who Benefits Most from Rural Electrification? Evidence from India." *Energy Journal* 35(2): 75-96.

Kowalski, Amanda E. 2016. "Doing More When You're Running LATE: Applying Marginal Treatment Effect Methods to Experiments."

Ligon, Ethan. 2015. "Estimating Household Neediness from Disaggregated Expenditures."

Lipscomb, Molly, Mobarak, Ahmed Mushfiq, and Tania Barham. 2013. "Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil." *American Economic Journal: Applied Economics* 5(2): 200-231.

Lee, Kenneth, Eric Brewer, Carson Christiano, Francis Meyo, Edward Miguel, Matthew Podolsky, Javier Rosa, and Catherine Wolfram. 2016. "Electrification for "Under Grid" Households in Rural Kenya." *Development Engineering* 1: 26-35.

Lee, Kenneth, Edward Miguel, and Catherine Wolfram. 2016. "Appliance Ownership and Aspirations among Electric Grid and Home Solar Households in Rural Kenya." *American Economic Review: Papers & Proceedings* 106(5): 89-94.

Lee, Kenneth, Edward Miguel, and Catherine Wolfram. 2018. "Experimental Evidence on the Economics of Rural Electrification."

McKenzie, David. 2012. "Beyond baseline and follow-up: The case for more T in experiments." *Journal of Development Economics* 99(2): 210-221.

Table 1. Primary outcomes

ID	Outcome	Unit	Type	Description	Ref.	R1	R2
P.1	Grid connected	HH	Indicator	Indicator for main household connection	1.1	x	x
P.2	Grid electricity spending	HH	Total	Estimated prepaid top-up last month or amount of last postpaid bill	1.7	x	x
P.3a	Employed or own business – Household	HH	Proportion	Proportion of household members (18 and over) currently employed or running their own business	4.5	x	x
P.3b	Total household earnings	HH	Value	Sum of earnings for all household members that are employed or running their own business, as well as agricultural earnings.	4.12	–	x
P.4	Total hours worked	Resp.	Total	Total hours worked in agriculture, self-employment, employment, and household chores in last 7 days	4.11	x	x
P.5	Total asset value	HH	Estimated value	Estimated value of savings, livestock, electrical appliances, and other assets	5.6	x	x
P.6	Annual consumption	HH	Value	Estimated value of annual consumption of 23 goods	6.2	x	x
P.7	Recent symptoms index	Resp.	Index	Index of symptoms experienced by the respondent over the past 4 weeks	7.3	x	x
P.8	Life satisfaction	Resp.	Scale	Life satisfaction based on a scale of 1 to 10	7.8	x	x
P.9a	Average test score	Child	Z-score	Average of English reading test result and Math test result	8.3	x	–
P.9b	Average exam score	Child	Z-score	Average result on KCPE examination (taken during follow-up period)		–	x
P.10	Political and social awareness index	Resp.	Index	Index capturing the extent to which the respondent correctly answered a series of questions about current events	9.4	x	–
P.11	Crime index	Resp.	Index	Index of crime experienced by the respondent over the last 12 months	11.2	–	x

Table 2. Mean effect indices

	Outcome	Unit	Type	Description	Ref.	R1	R2
G.1	Economic index	HH	Index	Index of primary economic outcomes	P3-P6	x	x
G.2	Non-economic index	HH	Index	Index of primary non-economic outcomes	P7-P11	x	x

Table 3. Energy consumption major outcomes

ID	Outcome	Unit	Type	Component(s)	R1	R2	Survey no. (R1, R2)
1.1	Grid connected	HH	Indicator	Indicator for main household connection	X	X	F1a
1.2	Electric lighting	HH	Indicator	Indicator for electricity as main source of lighting	X	X	F1b
1.3	Lighting usage	HH	Total	Hours of lighting used (past 24 hours)	X	X	F18
1.4	Installation	HH	Total	Number of electrical outlets available	X	X	F6b
				Number of lighting sockets available	X	X	F6c
				Number of power strips in use	X	X	F6e
1.5	Appliances owned	HH	Total	Number of "high-wattage" appliances owned[1]	X	X	F19a to F19c
1.6	Appliances desired	HH	Total	Number of "high-wattage" appliances desired	X	X	F19d to F19g
1.7	Grid electricity spending	HH	Total	Estimated prepaid top-up last month	X	X	F7a to F7e, F5h
				Amount of last postpaid bill	X	X	F8a to F8c, F5h
1.8	Kerosene spending	HH	Total	Kerosene spending last month[2]	X	X	F11
1.9	Other energy sources spending[3]	HH	Total	Solar power spending last month	X	X	F13d, F14d
				Battery spending last month	X	X	F15b, F15c
				Generator spending last month	X	X	F16c
				Purchased firewood spending last month	X	X	F17a
				Charcoal spending last month	X	X	F17b
				LPG spending last month	X	X	F17c
				Sawdust spending last month	X	X	F17d
				Mobile phone charging last month	X	X	F17h
				Other spending last month	X	X	F17e to F17g, F17i
1.10	Total energy spending	HH	Total	Total spending last month on grid electricity, kerosene, and other energy sources	X	X	See 1.7, 1.8, and 1.9 above

[1] In general, we follow Lee, Miguel, and Wolfram (2016a) in the definition of high and low wattage appliances. For instance, there we define mobile phones and radios as "low-wattage" appliances.
[2] For several energy spending categories (including kerosene), we recorded how much the household spent over the past seven days. In these cases, we will estimate spending over the past month by multiplying the weekly amount by a factor of approximately 4.3.
[3] This outcome will include all other energy-related expenditures recorded in the household survey, beyond grid electricity and kerosene.

1.11	Home solar usage	HH	Indicator	Indicator for usage of solar lantern or solar home system	x	x	F12a
1.12	Power sharing	HH	Indicator	Indicator for household is sharing its electricity connection (e.g., electricity connection shared with a minor household or a neighboring household)	x	x	A2 (*R1 only*), F5b, F5i, F5j, A18b (*R2 only*)

Table 4. Household structure major outcomes

ID	Outcome	Unit	Type	Component(s)	R1	R2	Survey no. (R1, R2)
2.1	Household size	HH	Total	Total number of household members	x	x	Section A
2.2	Inhabited location	HH	Indicator	Baseline structure currently inhabited	x	x	Staff records
2.3	Household stayed	HH	Indicator	Household did not move to a new location	x	x	Staff records, AA9
2.4	Members living elsewhere	HH	Total	Household members documented at baseline that are now living elsewhere	x	x	Section A
2.5	Fertility	Resp.	Total	Number of times respondent (or sexual partner) has been pregnant since January 2014	x	x	sH3_3num_m, sH3_3num_f
2.6	Local social interactions	Resp.	Total	Number of times (over past week) neighboring respondents visited household and respondent visited neighboring households	x	-	Section K

Table 5. Time use major outcomes

ID	Outcome	Unit	Type	Component(s)	R1	R2	Survey no. (R1, R2)
3.1	Hours sleeping	Resp.	Hours	Sleeping (code 1)	x	–	L1 to L48
3.2	Hours studying	Resp.	Hours	Playing with children or helping with homework (code 13)	x	–	L1 to L48
				Studying or attending class (code 16) *Note: All codes representing "studying" in survey*	x	–	
3.3	Hours working	Resp.	Hours	Light farm work (code 22)	x	–	L1 to L48
				Heavy farm work (code 23)	x	–	
				Fishing or hunting (code 24)	x	–	
				Office/desk work (code 25)	x	–	
				Light manual work (code 26)	x	–	
				Heavy manual work (code 27)	x	–	
				Other (work and travel) (code 32) *Note: All codes representing "work" in survey*	x	–	
3.4	Hours doing chores	Resp.	Hours	Cooking or preparing food (code 7)	x	–	L1 to L48
				Shopping for family (code 8)	x	–	
				Cleaning, dusting, sweeping, washing dishes or clothes, ironing, or doing other household chores (code 9)	x	–	
				Taking care of others, such as bathing, feeding, or looking after children, the sick, or the elderly (code 12)	x	–	
				Fetching water or firewood (code 10)	x	–	
				Repairs in or around the home (code 11)	x	–	
				Improving land or buildings (code 28) *Note: All codes representing "chores" in survey*	x	–	
3.5	Hours enjoying leisure	Resp.	Hours	Rest, watching TV, listening to the radio, reading a book, watching a movie, watching sports, or sewing (code 6)	x	–	L1 to L48
				Visiting or entertaining friends (code 14)	x	–	
				Playing sports (code 17)	x	–	
				Spending time with spouse or partner (code 18) *Note: All codes representing "leisure" in survey*	x	–	

Table 6. Productivity major outcomes

ID	Outcome	Unit	Type	Component(s)	R1	R2	Survey no. (R1, R2)
4.1	Agriculture – Land use	HH	Proportion	Proportion of total land used for agricultural activities	x	–	C4a, C4b, D1c
4.2	Irrigation	HH	Indicator	Household used irrigation in last 12 months	x	x	D2e
4.3	Agriculture – Monthly revenue	HH	Total	Revenue from selling crops	x	x	D4a
				Revenue from selling livestock or livestock products	x	x	D4c
				Revenue from selling poultry or poultry products	x	x	D4e
				Revenue from selling fish	x	x	D4g
				Revenue from selling other agricultural produce *Note: Household revenue over past month*	x	x	D4i
4.4	Agriculture – Hours worked	Resp.	Total	Hours worked in agriculture in last 7 days	x	x	D3a
4.5	Employed or own business - Household	HH	Proportion	Proportion of household members (18 and over) currently employed or running their own business	x	x	A8
4.6	Business at household	HH	Indicator	Business operated out of household compound	x	–	sE1_15cdescpremise, sE1_51otherbus
4.7	Employed or own business – Individual	Resp.	Indicator	Currently self-employed, running a business, employed, or working for pay	x	x	sE1_1selfemp, sE2_1employed
4.8	Employed or own business – Individual monthly compensation	Resp.	Total	Monthly compensation, sum of last month compensation across all jobs and businesses	x	x	sE2_11, sE1_9aprofit, sE1_56profit
4.9	Employed or own business – Individual hours worked	Resp.	Total	Hours worked in self-employment in last 7 days	x	x	sE1_5wrkhrs
				Hours worked in employment in last 7 days	x	x	sE2_7hours_1
4.10	Household chores – Individual hours worked	Resp.	Total	Hours spent doing household chores in last 7 days	x	x	sL_49hhchores
4.11	Total hours worked	Resp.	Total	Total hours worked in agriculture, self-employment, employment, and household chores in last 7 days	x	x	See 4.3, 4.8, and 4.9 above
4.12	Total household earnings	HH	Value	Sum of earnings for all household members that are employed or running their own business, as well as agricultural earnings. (last 30 days)	–	x	A8, D4 (*R2 only*)

Table 7. Wealth major outcomes

ID	Outcome	Unit	Type	Component(s)	R1	R2	Survey no. (R1, R2)
5.1	Savings	Resp.	Total	Savings in mobile bank account	x	x	G2a
				Savings in SACCO, merry-go-round, or ROSCA	x	x	G2b
				Savings in formal bank account	x	x	G2c
5.2	Housing quality	HH	Index	Indicator for high-quality floors	x	x	C1a
				Indicator for high-quality roof	x	x	C1b
				Indicator for high-quality walls	x	x	C1c
5.3	Value of livestock assets	HH	Estimated value	Value of chickens owned	x	x	C8a
				Value of cattle owned	x	x	C8b
				Value of goats owned	x	x	C8c
				Value of pigs owned	x	x	C8d
				Value of sheep owned	x	x	C8e
5.4	Value of appliance assets	HH	Estimated value	Value of listed electrical appliances	x	x	F19a to F19c
5.5	Value of other assets	HH	Estimated value	Value of beds owned	x	x	C7a
				Value of bednets owned	x	x	C7b
				Value of kerosene stoves owned	x	x	C7c
				Value of kerosene lamps owned	x	x	C7d
				Value of hoes owned	x	x	C7e
				Value of bicycles owned	x	x	C7f
				Value of motorcycles owned	x	x	C7g
				Value of cars or trucks owned	x	x	C7h
				Value of sofa piece seats owned	x	x	C7i
5.6	Total asset value	HH	Estimated value	Estimated value of savings, livestock, electrical appliances, and other assets	x	x	See 5.1, 5.3, 5.4, and 5.5 above

Table 8. Consumption major outcomes

ID	Outcome	Unit	Type	Component(s)	R1	R2	Survey no. (R1, R2)
6.1	Neediness index	HH	Index	Consumption of each of 23 goods over past twelve months, constructed according to the measure in Ligon (2015)	X	–	M5, M7, M8
6.2	Annual consumption	HH	Value	Estimated value of annual consumption of 23 goods[4]	X	X	M5, M7, M8 (for Round 1) sm_1staple, sm_2veg, sm_3meat, sm_4fruit, sm_5other (for Round 2)
6.3a	Consumption diversity	HH	Index	Indicators for whether household has consumed each of 23 goods over the past twelve months	X	–	M1
6.3b	Consumption diversity	HH	Index	Indicators for whether household has consumed each of 23 goods over the past week	–	X	M1
6.4	Meals	Resp.	Total	Total number of meals eaten yesterday	X	X	sH1_1meals
6.5	Protein meals	Resp.	Total	Total number of meals eaten yesterday including meat or fish	X	X	sH1_2ameat

[4] Note that for Round 2, we will estimate annual consumption based on how the respondent reports consumption from the past seven days.

Table 9. Health and wellbeing major outcomes

ID	Outcome	Unit	Type	Component(s)	R1	R2	Survey no. (R1, R2)
7.1	Respiratory illness index	Resp.	Index	Persistent cough	X	–	sH1_7bcough
				Asthma/breathlessness at night *Note: Experienced over past 4 weeks*	X	–	sH1_7sasthma
7.2	Respiratory illness index - Child	Child	Index	Frequent cough	X	–	T3.5
				Itchy or stinging eyes	X	–	
				Sore throat	X	–	
				Runny nose	X	–	
				Asthma or breathlessness *Note: Experienced over past 7 days*	X	–	
7.3	Recent symptoms index	Resp.	Index	Fever	X	–	sH1_7afever
				Persistent cough	X	–	sH1_7bcough
				Persistent tiredness	X	–	sH1_7ctired
				Stomach pain	X	–	sH1_7dstomach
				Blood in stool	X	–	sH1_7fstool
				Rapid weight loss	X	–	sH1_7gweightloss
				Frequent diarrhea	X	–	sH1_7hdiarrhoea
				Skin rash or irritation	X	–	sH1_7iskin
				Open sores/boils	X	–	sH1_7jboils
				Difficulty swallowing	X	–	sH1_7kswallow
				Sores or ulcers on the genitals	X	–	sH1_7pgenitalsore
				Asthma/breathlessness at night	X	–	sH1_7sasthma
				Frequent and excessive urination	X	–	sH1_7tfrequrine
				Constant thirst/increased drinking of fluids	X	–	sH1_7uthirst
				Unusual discharge from the tip of penis *(for men only)*	X	–	sH1_7wdischarge
				Other symptoms	X	–	sH1_7xother
				Note: All symptoms experienced over past 4 weeks			
7.4	Recent illnesses index	Resp.	Index	Worms	X	–	sH1_7eworms

				Malaria	x	–	sH1_7mmalaria
				Typhoid	x	–	sH1_7ntyphoid
				Tuberculosis	x	–	sH1_7otb
				Diabetes	x	–	sH1_7vdiabetes
				Cholera	x	–	sH1_7qcholera
				Yellow fever	x	–	sH1_7ryellow
				Note: All illnesses experienced over past 4 weeks	x	–	
7.5	Recent illnesses index - Child	Child	Index	Malaria Fever Typhoid *Note: All symptoms experienced over past 7 days*	x x x	– – –	T3.5
7.6	Subjective health	Resp.	Indicator	Self-described health is either "good" or "very good"	x	–	sH1_13healthgd
7.7	Subjective health - Child	Child	Indicator	Self-described health is either "good" or "very good"	x	–	T3.4
7.8	Life satisfaction	Resp.	Scale	Life satisfaction based on a scale of 1 to 10	x	–	J9b

Table 10. Education major outcomes

ID	Outcome	Unit	Type	Component(s)	R1	R2	Survey no. (R1, R2)
8.1	English score	Child	Z-score[5]	English reading test result	x	–	T1
8.2	Math score	Child	Z-score	Math test result	x	–	T2
8.3	Average test score	Child	Z-score	Average of English reading test result and Math test result	x	–	T1, T2
8.4	Study hours - Total	Child	Total	Self-reported hours spent studying during the day	x	–	T3.1
				Self-reported hours spent studying during the night	x	–	T3.2
8.5	Study hours - Night	Child	Total	Self-reported hours spent studying during the night	x	–	T3.2
8.6	Attendance index	Child	Index	Fully completed first week of school last term	x	–	B2b
				Fully completed last week of school last term	x	–	B2c
				Completed end of term exams last term	x	–	B2d
				Fully completed first week of school this term	x	–	B2e
8.7	Grades	Child	Score	Marks (scaled out of 100) earned last term	x	–	B2f
8.8	Ambitions	Child	Indicator	Student planning to attend post-secondary education	x	–	T3.7
8.9	KCPE	Child	Indicator	Child took KCPE exam	–	x	A100
8.10	KCPE score	Child	Z-score	KCPE score	–	x	A100
8.11	KCPE age	Child	Age	Age of child when took KCPE exam	–	x	Section A

[5] We will create Z-scores by subtracting the mean and dividing by the standard deviation in the control group within our own sample, using age-gender groups.

Table 11. Social and political attitudes major outcomes

ID	Outcome	Unit	Type	Details	R1	R2	Survey no. (R1, R2)
9.1	Radio	Resp.	Total	Days in the past week respondent listened to the radio	X	X	J2a
9.2	Television	Resp.	Total	Days in the past week respondent watched television	X	X	J2c
9.3	Internet	Resp.	Total	Days in the past week respondent used the internet	X	X	J2d
9.4	Political and social awareness index	Resp.	Index	Knows date of next election	X	–	J1a
				Knows name of the president of Tanzania	X	–	J1b
				Knows name of the president of Burundi	X	–	J1c
				Knows name of a candidate in the 2016 U.S. presidential election	X	–	J1d
				Knows name of the CEO of Safaricom	X	–	J1e
				Knows name of the Managing Director of Kenya Power	X	–	J1f
				Knows the intended recipients of the Kenyan national government's Free Laptop program	X	–	J1i
				Knows who was responsible for the 2015 terrorist attacks at Garissa University	X	–	J1j
				Knows which team won the 2015-2016 English Premier League	X	–	J1g
				Knows who sings the pop song "Sura Yako" *Note: These are all binary variables*	X	–	J1h
9.5	Approval of national government index	Resp.	Index	Trusts national government	X	–	J5g
				Uhuru Kenyatta is doing a good job as president	X	–	J7a
				Government is doing a good job fighting terrorism	X	–	J7b
				Government corruption is <u>not</u> a problem in Kenya	X	–	J7d
				Government is doing a good job ensuring that electricity is provided in Kenya *Note: Binary variable indicating "agree" or "strongly agree"*	X	–	J7g
9.6	Gender equality index	Resp.	Index	It is acceptable for a woman to be a bus driver	X	–	J6a
				Important decisions of the family should <u>not only</u> be made by the man of the family	X	–	J6b
				If the wife is working outside the home, the husband should help her with household chores	X	–	J6c

				Women should have more opportunities to become political leaders *Note: Binary variable indicating "agree" or "strongly agree"*	X	–	J6d
9.7	Ethnic identity index	Resp.	Index	Ethnic identity is "important" or "very important" in respondent's life	X	–	J4e
				Indicator for belongs first to ethnic group (over other dimensions of identity)	X	–	J4f
9.8	Religiosity index	Resp.	Index	Religious identity is "important" or "very important" in respondent's life	X	–	J4d
				Indicator for belongs first to religious group (over other dimensions of identity)	X	–	J4f
				Attends church/mosque regularly	X	–	J4a
				Attended church/mosque last week	X	–	J4b
9.9	Social trust index	Resp.	Index	Trusts people, in general	X	–	J5a
				Trusts members of their own ethnic group	X	–	J5b
				Trusts members of other ethnic groups	X	–	J5c
				Trusts members of their own religion	X	–	J5d
				Trust members of other religions *Note: Indicator for "can be trusted" or "can be somewhat trusted"*	X	–	J5e

Table 12. Community primary outcomes

ID	Outcome	Unit	Type	Details	R1	R2	Survey no. (R1, R2)
10.1	Comm. electrification rate	Com.	Proportion	Estimated community electrification rate	X	X	See Section 2.3
10.2	Comm. electricity reliability index	Com.	Index	Proportion of connected households reporting power blackouts in past 7 days	X	X	F10c, F10d
				Proportion of connected households reporting regular blackouts	X	X	F10e
10.3	Value of assets inequality	Com.	Index	Gini coefficient capturing within-community dispersion in total asset value	X	X	See 5.6 above
10.4	Education inequality	Com.	Index	Gini coefficient capturing within-community dispersion in student test score results (we use tests administered by the research team for Round 1 and KCPE scores for Round 2)	X	X	T1, T2 For R2, see 8.10 above
10.5	Consumption inequality	Com.	Index	Gini coefficient capturing within-community dispersion in total consumption of 23 consumption goods	X	–	M5, M7, M8
10.6	Gini coefficient for weekly consumption	Com.	Index	Gini coefficient capturing within-community dispersion in total consumption of 23 consumption goods	–	X	Section M
10.7	Perceived income inequality	Com.	Proportion	Proportion of respondents agreeing with statement that economic inequality is a problem in this village	X	–	J7e
10.8	Gini coefficient for total household earnings	Com.	Index	Gini coefficient capturing within-community dispersion in total household earnings	–	X	A8, D4
10.9	Voting	Com.	Proportion	Proportion of community that voted for Uhuru Kenyatta in the August 2017 general elections	–	X	Voting data from August 2017 national elections

Table 13. Safety and crime major outcomes

ID	Outcome	Unit	Type	Details	R1	R2	Survey no. (R1, R2)
11.1	Safety	Resp.	Indicator	Area is described as "very secure" or "secure"	–	X	J10
11.2	Crime	Resp.	Index	Livestock stolen in last 12 months Household items stolen in last 12 months Cash stolen in last 12 months Assaulted without weapon in last 12 months Assaulted with weapon in last 12 months Victim of arson in last 12 months Victim of witchcraft in last 12 months Other crime in last 12 months	–	X	J11 J12 J13 J14 J15 J16 J17 J18