MATHEMATICA Policy Research

REPORT

Evaluation of the Tanzania Energy Sector Project: Final Update of Design Report

July 28, 2015

Duncan Chaplin
Arif Mamun
Candace Miller
Ali Protik
John Schurrer

Submitted to:
Millennium Challenge Corporation
875 15th Street, NW
Washington, DC 20005
Project Officer: Jenny Heintz and Shreena Patel
Contract Number: MCC-10-114-CON-20-TO05

Submitted by:
Mathematica Policy Research
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director: Arif Mamun and Duncan Chaplin
Reference Number: 06919.415

This page has been left blank for double-sided copying.

ACKNOWLEDGMENTS

We would like to thank a number of people whose efforts directly or indirectly culminated in the completion of this evaluation design report. We are especially grateful to our monitoring and evaluation colleagues at the Millennium Challenge Corporation (MCC): John "Jack" Molyneaux, Jenny Heintz, Jennifer Sturdy, Shreena Patel, Laura Rudert, and Anne Rothbaum Pizer. Additionally, we thank Scott Alexander, Karl Fickenscher, and Matthew Kavanagh, each of whom is or was with MCC's Resident Country Mission in Tanzania, as well as Himesh Dhungel and Kate Iovanna from the MCC energy sector team. Their input at all stages of the project has been invaluable.

In addition, we appreciate the cooperation and support from our colleagues at the Millennium Challenge Account—Tanzania (MCA-T), including Bernard Mchomvu, Ahmed Rashid, Paschal Assey, Issac Chanji, Chedaiwe Luhindi, Joseph Hayuni, Salum Ramadhani, Peter Kigadye, Florence Gwang'ombe, Reginald Ndindagi, Athanas Aloisi, William Christian, and Deborah Sungusia.

We would like to recognize the other members of the Tanzania energy sector evaluation team, including Mathematica's Kathy Buek for directing our efforts in overseeing the baseline data collection, and Divya Vohra and Anca Dumitrescu for their work on the follow-up data collection. Denzel Hankinson, Josh Finn, Lauren Pierce, and Brendan Larkin-Connolly of DHInfrastructure provided valuable input at various stages of the evaluation. Leonard Sibomana, our consultant in Tanzania, did excellent work in collecting the sampling frame information for the baseline community survey and in overseeing the baseline data collection process.

This report also benefited greatly from discussions with our colleagues at Mathematica. These include Ken Fortson, Phil Gleason, Peter Schochet, Frank Potter, Nancy Murray, Steve Glazerman, and Lesley Hildebrand. We would also like to thank Kristine Johnston, Hannah Burak, Elizabeth Kelley, Nicholas Redel, Ebo Dawson-Andoh, Ji-Hyeun Kwon-Min, and Miles Watkins for excellent research assistance; Xiaofan Sun, Yuhong Zheng, and Cathy Lu for programming support; Betty Teller for editing the report; and Colleen Fitts for formatting the report. Finally, Ryan Collins, Lindsay Eckhaus, and Janine O'Donnell provided invaluable assistance in tracking and managing the project staffing and resources.

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURES

ACRONYMS

DID	Difference-in-differences
EDI	Economic Development Initiatives
FS	Financing scheme (the customer-connection financing scheme initiative)
GDP	Gross domestic product
GPS	Global positioning unit
IGA	Income-generating activity
kg	Kilogram
MCA-T	Millennium Challenge Account—Tanzania
MCC	Millennium Challenge Corporation
MDI	Minimum detectable impacts
MoF	Ministry of Finance
NBS	National Bureau of Statistics
NRECA	NRECA International
PSU	Primary sampling unit
T&D	Transmission and distribution
TANESCO	Tanzania Electric Supply Company
TZS	Tanzanian shilling [1 USD = 1,577 TZS]
USD	U.S. dollar

This page has been left blank for double-sided copying.

The United States and many other countries are showing a growing commitment to providing the African continent with more access to electric power. Commitments to the Power Africa initiative, estimated at around $18 billion US, are just one illustration of this resolve (USAID 2014). With only a small fraction of its population having access to electricity, Tanzania provides a good example of the need for these efforts. In recognition of the potential for the Tanzanian government to help address this need, the Millennium Challenge Corporation (MCC) funded a $207 million energy sector project in Tanzania in 2008. This project was implemented by the Millennium Challenge Account–Tanzania (MCA-T), a part of the Tanzanian government.

MCC contracted with Mathematica Policy Research to carry out evaluations of the energy sector project, including four major components of the project: (1) the distribution systems rehabilitation and extension activity, (2) a customer-connection financing scheme (FS) initiative to facilitate lower-cost electricity connections in selected areas, (3) the Kigoma solar power activity, and (4) the Zanzibar interconnector activity or "cable" activity. The distribution systems rehabilitation and extension activity, also known as the transmission and distribution (T&D) activity, was designed to provide new electricity lines to around 350 communities spread throughout seven regions of mainland Tanzania, the FS initiative offered low-cost connections to about 5,800 households in 29 of these 350 communities, and the Kigoma solar activity was designed to provide solar power access to a broad variety of potential beneficiaries in the Kigoma region of Tanzania, including schools, health facilities, businesses, and fisherman, as well as to develop a market for solar systems for households. The cable activity was designed to improve access to electricity on the Unguja Island of Zanzibar by adding a new submarine cable to transmit grid electricity from the mainland and by upgrading various other components of the Zanzibar electrical grid. The evaluation of these components is designed to address a number of research questions. Broadly speaking, the questions can be divided into two complementary categories:

- **Impact evaluation**. What are the impacts of the project components on key outcomes related to energy use, health, education, employment, and income? Are there unintended consequences? Would a less rigorous evaluation produce similar results? How do impacts vary by subgroup? What lessons can be learned from the impact findings?
- **Performance evaluation.** How well were these components of the program implemented? What challenges were encountered? What lessons can be learned from the implementation of the program? Do the performance results help us interpret the impact findings?

In order to estimate impacts of the T&D activity and the FS initiative, we developed rigorous evaluation designs. We are using a quasi-experimental **difference-in-differences (DID**) **design with a matched comparison group** to estimate the impacts of T&D line extensions. For estimating the impacts of the FS initiative, we are using a **random assignment evaluation design**, which is considered to be the gold standard for impact evaluations. The way in which the remaining two components were implemented limited us to less rigorous **pre-post** designs for those components. Another evaluator collected the baseline data for the Kigoma solar activity evaluation. They obtained pre-intervention data for some beneficiaries but only retrospective data for others. In addition, while they did obtain data for a few comparison units those units were not chosen using a clearly described matching process. For the now-completed cable activity evaluation, we conducted a case study of a few large hotels in Zanzibar. These hotels employed a large number of people in Zanzibar but, even so, we are limited in our ability to generalize to the full set of potential beneficiaries, which includes all other businesses in Zanzibar, as well as households. Although these less rigorous evaluation designs have limitations, they can provide valuable insights regarding potential benefits of the activities. In this report, we describe our evaluation designs for all four energy sector program components.

I. INTRODUCTION

Access to reliable, high quality electricity can be a key driver of economic growth and household well-being (Barnes 1988; World Bank 2008). In Tanzania, only about 12 percent of all households in the mainland had access to the national electricity grid in 2007—and the rate was just 2.5 percent in rural areas (NBS 2009). In addition to the low level of electrification in the country, the power available is subject to frequent surges and interruptions in service for many customers. With a gross domestic product (GDP) per person of only US$675 per person in 2013 (World Bank 2014), Tanzania is one of the poorest countries in the world, ranking in the bottom 15 percent of countries in a recent report (United Nations 2014). Nearly 33 percent of the population in mainland Tanzania and 49 percent of the Zanzibar population live below the poverty line, as determined by Tanzania's Ministry of Finance (MoF 2009; Zanzibar MoF 2009).

In an effort to promote economic growth and reduce poverty in Tanzania, the Millennium Challenge Corporation (MCC) funded an energy sector project that was implemented by the Millennium Challenge Account–Tanzania (MCA-T). The project has a number of key components, the largest of which is the distribution systems rehabilitation and extension activity, also known as the transmission and distribution (T&D) activity. Other component of the project include a customer-connection financing scheme initiative to facilitate lower-cost electricity connections in selected areas (hereinafter, financing scheme initiative or FS initiative), promotion of solar power systems in the Kigoma region of mainland Tanzania (Kigoma solar), and installation of a new submarine cable connecting Zanzibar's Unguja Island to the mainland along with rehabilitation of various parts of the Zanzibar grid (the Zanzibar interconnector activity, or cable activity). Together, these activities were designed to increase the availability of reliable and high quality electricity to people in mainland Tanzania and Zanzibar.

MCC contracted with Mathematica Policy Research to carry out evaluations of the four components of the energy sector project described above. These evaluations are designed to enable MCC to understand more fully how these components of the project affected the well-being of the target populations.

This design report is based in part on an earlier design report (Chaplin et al. 2011), which was itself based on an even earlier design memorandum. This current (and final) design report also incorporates updates to our design described in our baseline report (Chaplin et al. 2012). The major changes since the 2012 baseline report are as follows:

- Two parts of our evaluation have been completed—the evaluation of the cable activity and a qualitative report that focused on the T&D activity and FS initiative.
- The 2012 baseline report also described plans to do an interim impact report using data on connections per community provided by the local electricity utility. After extensive review, it was decided that the connections data were not reliable enough to do an interim impact analysis. That decision is described in more detail in a separate memo (Chaplin et al. 2015).
- We are adding a description of our plans for an evaluation of the Kigoma solar activity.
- We are also adding a more detailed description of how we plan to deal with the issue of migration of households.

- We are clarifying which outcomes will be the focus of the T&D evaluation.

The updated evaluation design report serves two purposes: it conveys the key evaluation design decisions, and also allows us to indicate that these decisions are not influenced by data on outcomes of interest. At the time of preparing the updated report the follow-up community survey is being implemented, and we have started to look at some of the community survey data, which could create a cause for concern. However, our analyses of those data have been focused almost entirely on checking them for internal consistency. The only outcomes we have looked at for the intervention and comparison groups separately are connection rates and the locations of the new MCC and non-MCC funded poles and we are not making any changes in the planned analysis of those outcomes. We have not looked at any data by whether the community benefitted from the FS initiative. Also, with the exception of the changes listed above, we are making no additional changes to the rigorous parts of the evaluation design (for estimating impacts of the T&D activity and the FS initiative). Those were described in Chaplin et al. (2011) and Chaplin et al (2012), long before we had access to any follow-up data. This report simply repeats the descriptions made there.

We begin this report with an overview of the Tanzania energy sector project, a brief review of the literature on impacts of electrification, and a discussion of the conceptual framework that guides the Tanzania energy sector evaluation. In Chapter II, we present the evaluation design for the T&D activity and the financing scheme initiative, the baseline survey data collection and sampling for these surveys, and statistical power for the impact analysis. In Chapter III, we present our plans for the Kigoma solar activity. In Chapter IV, we discuss changes in our plans for addressing migration in the evaluation of the T&D activity and the FS initiative. In Chapter V, we discuss the Zanzibar cable evaluation. Finally, in Chapter VI, we discuss our timeline and next steps. In the appendix, we present a technical discussion on sampling and matching weights

A. Overview of the energy sector project

Tanzania is one of a handful of nations awarded a compact from MCC. At about $698 million, the Tanzania compact is the largest MCC compact to date. In order to effectively manage the work of this compact, the Tanzanian government created MCA-T, which is now implementing the project activities with oversight from MCC. To address infrastructure constraints to economic growth and poverty reduction in the country, MCA-T is using the MCC compact to fund projects in three sectors: roads, water, and energy. In particular, MCC is investing $207.2 million in four components of the energy sector project:[1]

- **Distribution systems rehabilitation and extension activity (T&D activity**). This activity involves rehabilitation of existing electricity transmission and distribution networks[2] as well as construction of new lines in Dodoma, Iringa, Kigoma, Mbeya, Morogoro, Mwanza, and

[1] For more details on the energy sector project activities, see Annex I in the Tanzania Millennium Challenge Compact (MCC 2008), available at http://www.mcc.gov/documents/agreements/compact-tanzania.pdf.

[2] Even though transmission lines usually refer to electricity lines of 66 kilovolts or higher capacity, and all electricity lines built under the Tanzania energy project were 33/11 kilovolts or lower capacity, the lines built under the project have been referred to as transmission and distribution lines.

Tanga—regions identified as being high priority for investment in electricity.[3] The $126.2 million being invested in the T&D activity represents more than three-fifths of MCC's total investment in the energy sector project.

- **The financing scheme initiative (FS initiative**). MCC and MCA-T are concerned that many households will not be able to afford to connect to the new lines created by the T&D activity. Consequently, they are funding a separate but closely related financing scheme initiative to facilitate 5,800 lower-cost connections for households in 29 communities selected from the 350 or so that received the new T&D lines. A communications campaign will be carried out as part of the initiative, to inform households about the low-cost connection offer that will be available for a limited time and on a first-come, first-served basis.
- **Zanzibar interconnector activity (cable activity).** This activity is designed to improve the quality and reliability of the electricity to Unguja Island in Zanzibar by installing a new submarine cable from the mainland, upgrading substations at either end of the cable, and installing new overhead cables on both the mainland and Unguja Island. With about $68 million being invested in it, the cable activity represents about one-third of MCC's total investment in the energy sector project.
- **Kigoma solar power activity (solar activity).** This activity involves installing solar modules and other solar electric systems in 45 schools, 130 health facilities, 45 markets, and 90 fishing boats, as well as development of a market for solar systems for households, all in the Kigoma region of Tanzania. Almost all direct beneficiaries of the activity are slated to receive photovoltaic power.

Through these investments in the energy sector, MCC aims to help Tanzania take fuller advantage of its economic growth potential and ultimately improve the well-being of its people. Mathematica's evaluation of the energy sector project will help assess how successful these components of the energy sector project have been in achieving those goals. In the remainder of this chapter, however, we provide a conceptual framework for the overall energy sector project, along with a brief review of the empirical evidence that helps underscore the conceptual framework.

B. Evidence on impacts of expansion of grid electricity

There is limited rigorous evidence regarding the impacts of electrification, and much of what does exist focuses on impacts of rural electrification on poverty, education, health, and the environment. There is a dearth of rigorous research on the impact of electrification on peri-urban areas (that is, locations on the periphery of urban areas); such research would have been relevant for the Tanzania energy sector project, as many of the MCC-funded T&D lines are being built in urban/peri-urban areas, though many are also in rural areas. These urban/peri-urban areas may benefit more than rural areas since they are likely to have better infrastructural support for industrial and commercial development that can create higher-wage jobs than currently exist in these areas.

[3] The communities were selected from seven regions using definitions of regions supplied by TANESCO in 2011. Regional boundaries have changed and TANESCO regions do not always correspond with those defined for general administrative purposes, so a number of the communities covered by the T&D project are no longer located in the original seven regions.

The primary challenge in improving household well-being through electrification is the continued low rate of connection to the electricity grid among households even in communities that are covered by the electricity distribution network. This is particularly prevalent among poorer rural households across many countries in Africa (see, for example, ESMAP [2007b] for Senegal, Jacobson [2007] for Kenya, Ketlogetswe et al. [2007] for Botswana, Heltberg [2003] for South Africa and Ghana, and World Bank [2008]). The low connection rates are a result of relatively high connection costs and/or high tariffs that many households are unable to afford. The financing scheme initiative under the Tanzania energy sector project is expected to address this challenge in selected areas where new T&D lines are being extended, and provides an opportunity to assess the effects of electrification when a larger percentage of households are connected to the electric grid.

Evidence suggests that when households are connected to the electric grid, benefits accrue to them primarily through consumption of electricity for lighting, entertainment, and increased home and farm production. There are only a handful of rigorous evaluations in this regard. Their findings indicate that rural electricity reduces expenditures on lighting (Bernard and Torero [2009] in Ethiopia), increases time for household work, facilitates entry of women into the labor market (Dinkelman [2011] in South Africa), and increases farm income through irrigation (Khandker et al. [2009] in Vietnam). Although there is some evidence of benefits through increased economic activity and improved health and educational services at the community level, the benefits are smaller or less clear relative to the benefits that directly accrue to the household (Bernard and Torero 2009; Dinkelman 2011; World Bank 2008).

Considering the limited rigorous evidence on the impacts of electrification on household well-being and economic activities in developing country settings, the Tanzania T&D evaluation is expected to fill in some of the gaps in the literature. More specifically, the evaluation will provide estimates of short-term effects of electrification in rural, urban, and peri-urban areas in Tanzania. In the process, the evaluation is expected to generate critical information for policymakers in international development agencies as well as in Tanzania, and to provide input for future energy policy in the country.

C. Project logic and conceptual framework for the evaluation

MCC and MCA-T have developed a set of logic models for each activity under the energy sector project (MCA-T 2012). Mathematica consolidated the logic models in a conceptual framework, presented in Figure I.1, which guides our approach to the evaluation of the project activities. The boxes on the far left of the figure show the four energy sector activities. The box on the far right shows the ultimate objectives of the activities—increased economic growth, improved standard of living, and poverty reduction. The project activities are designed to achieve these objectives through their effects on access to electricity, which will be realized in the short term, and through subsequent effects on households, businesses, and communities, which will be realized in the intermediate and longer terms.

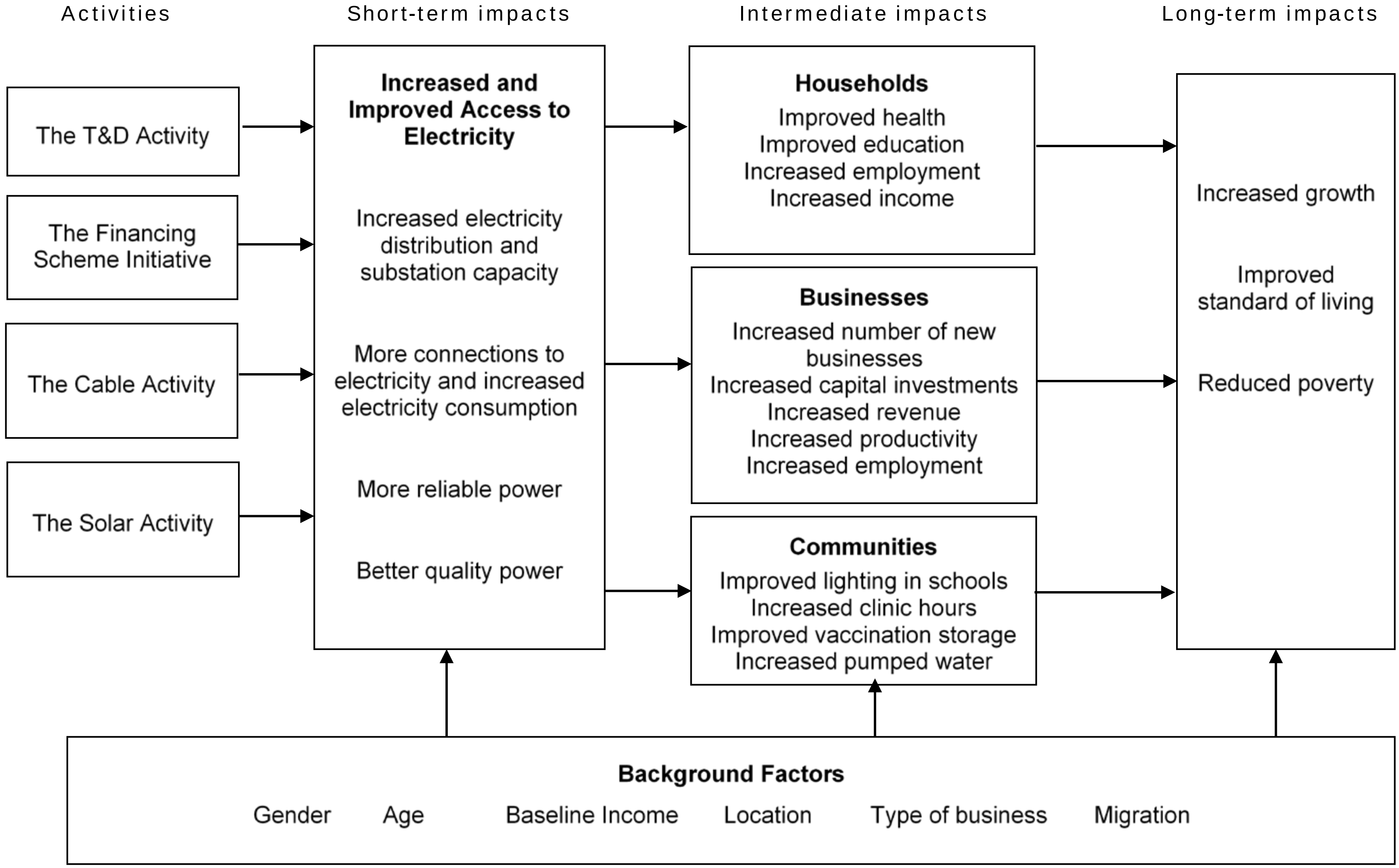
The energy project activities can affect access to electricity in several ways, as shown in the box in the second column of the conceptual framework. First, the successful implementation of the T&D activity is expected to increase the reach of the distribution networks and improve the substation capacity. Second, by expanding the distribution network and facilitating lower-cost connections, the T&D activity and the financing scheme initiative can increase the number of households, businesses, and community organizations (such as schools, health facilities, and water utilities) connected to the national grid, which would likely lead to increased use of electricity. Third, the provision of solar modules and other solar electric systems to different types of beneficiaries through the Kigoma solar activity is expected to increase access to electricity in some areas currently lacking the grid. Fourth, by installing a new submarine interconnector cable, the cable activity is designed to reduce the extent of service interruptions or outages, referred to as the “reliability” of electricity supply, in Zanzibar. Fifth, the installation and rehabilitation of electricity infrastructure in the T&D activity may reduce variations in voltage magnitude or harmonic distortions, referred to as the “quality” of electricity supply, which is expected to reduce equipment damage at the electric utilities and in homes and businesses.

These improvements in access to electricity can have important intermediate impacts on households, businesses, and communities, as presented in the third column of the conceptual framework. Electricity can help improve households’ economic opportunities by enabling household members to spend less time doing household chores during the day, consequently freeing up time to work for pay outside the home. It can also help households obtain valuable information on the market prices of goods and services, adverse weather conditions, and opportunities available to them, via radio and television programming and mobile phone communications. Electricity can improve health outcomes if it enables households to reduce use of certain types of fuel that are particularly likely to cause health problems, such as charcoal and wood. Finally, it can improve education outcomes by enabling students to spend more time reading after dark.

Electricity can also have important impacts for businesses. In particular, it can enable businesses to use many types of machinery that cannot be operated cost-effectively without electricity. Similarly, electricity can be used in important and cost-effective ways by facilities that serve entire communities, such as schools (which can benefit from electric lights), clinics (which can stay open for longer hours, use electricity for refrigeration, and use certain types of medical equipment), and water utilities (which can use electricity for pumps and cleaning equipment). For all of these types of uses, grid electricity from the new T&D lines funded by MCC can be far less expensive than electricity produced by the small generators commonly used by many businesses, schools, and health facilities that are operating away from the existing electric grid.

The box at the bottom of the framework shows background factors that may affect the short-term, intermediate, and long-term outcomes we are studying. It will be important to control for differences in these background factors when conducting our impact analyses. In addition, impacts of the activities may vary across different subgroups of the population. Women and children, for example, may benefit most from electricity in the house, since they spend more time there. Low-income households may benefit least if they cannot afford the connection fee or electric appliances. Benefits to businesses may depend on their use of electrical equipment. Communities may differ in the benefits they gain from electricity, depending on the number and type of public facilities they operate. Our evaluation will pay particular attention to differences by gender, as that is a strategic priority for MCC and MCA-T. Finally, migration may matter, especially if large numbers of new households migrate into communities that become electrified. Although we will not be able to rigorously estimate benefits for these households, we do plan to do a number of related analyses to help capture impacts that might be related to migration, as discussed in Chapter IV.

Figure I.1. Conceptual framework for the Tanzania energy sector project



II. T&D EVALUATION DESIGN, METHODS, AND DATA

In this chapter, we present the design, methods, and data sources for the evaluation of the T&D activity. The aim of the T&D evaluation is to assess the implementation successes and challenges, as well as to estimate the impacts of the T&D line extensions and the FS initiative. We discuss the central research questions the evaluation will address; key aspects of the impact estimation methods, sampling, and baseline surveys; and statistical power for identifying impacts of the T&D activity and FS initiative. We then describe the design for our now-completed qualitative analysis (Miller et al. 2015) and plans for an interim impacts analysis that we were forced to cancel due to poor data quality (Chaplin and Mamun 2015). A timeline for the intervention and evaluation is presented in Figure VI.1.

A. Evaluation questions

The T&D evaluation is designed to address a number of research questions that were selected in collaboration with MCC and MCA-T. It will answer the following impact and performance evaluation questions regarding the T&D activity:

1. Impact evaluation questions

- **Impacts on outcomes.** Does access to electricity lead to (1) increased connection to the national electric grid; (2) increased household income and better health and education outcomes; (3) increased business activity, including creation of new firms, capital investments, and greater levels of employment; and (4) improved community outcomes related to schools, hospitals, or water supply? If impacts are detected, what are the magnitudes of those impacts?
- **Unintended consequences**. Are there unintended impacts of the program (positive or negative)?
- **Benefits of a rigorous evaluation.** Does a rigorous evaluation design yield the same impact estimates as a simple pre-post design?
- **Subgroup analyses.** Do the impacts vary by gender, age, and income?
- **Lessons learned.** What are the implications of the evaluation findings for future electricity projects and long-term policymaking?

2. Performance evaluation questions

- **Implementation successes**. Were the interventions under the T&D activity implemented successfully? How well was the T&D activity implemented relative to its goals? How was the activity perceived by potential and actual beneficiaries? Was the activity sustained over time?
- **Challenges encountered.** What challenges were encountered in implementing the activity? How were the challenges addressed?
- **Lessons learned.** What are the lessons learned from the implementation of the activity?

The impact evaluation questions are addressed by our impact evaluation plans, below. We addressed the performance questions in our qualitative report, plans for which we provide below. The interim report would have provided additional information about connection rates over time. Although that report has been cancelled, we will still be able to estimate impacts on connection rates in 2015, using the community and household survey data collected for the final impact analysis.

B. Impact evaluation design for the T&D activity

In Table II.1, we summarize the technical approach for impact evaluation of the T&D line extensions and the FS initiative under the T&D activity.

Table II.1. Technical approach to impact evaluation: T&D line extension and FS initiative

Intervention	Evaluation methodology	Intervention/ treatment group	Comparison/ control group	Key outcomes
T&D line extension	Difference-in-differences (DID) method, which compares changes in outcomes over time between T&D intervention and matched comparison areas	Households, businesses, and communities in areas that received line extensions	Households, businesses, and communities in matched areas that did not receive new line extensions	Connection to, reliability, and quality of electricity Household income and expenditures Business energy expenditures and revenue Employment Health outcomes Child schooling attainment/ intensity of study Distribution of time and resources within the household by gender
Financing scheme (FS) initiative	Random assignment of areas either to a treatment or a control group; compare outcomes between these two groups at follow-up	Households in areas that received the T&D lines and the FS offers	Households in areas that received the T&D lines but did not receive the FS offers	

1. Matched comparison group evaluation design for T&D line extensions

We are using a difference-in-differences (DID) method with matched comparison group design to estimate the impacts of extending electricity lines to the new areas covered by the T&D activity. We will compare changes over time in outcomes for intervention communities in six regions of the country (that is, communities that will receive the line extensions) with changes in outcomes for comparison communities.[4,5] The households in the comparison communities were

[4] For the evaluation of the T&D activity, we refer to the areas receiving the line extensions as the "intervention group." A subset of that group will receive low-cost connections through the FS initiative. We refer to that subset as the "treatment group."

[5] The T&D activity is also being implemented in a seventh region—Kigoma. However, because Kigoma was not initially part of the T&D activity, no baseline surveys were conducted there; consequently the T&D evaluation will

chosen using propensity score matching so that they are similar to the households in the intervention communities based on various household characteristics, such as income, assets, consumption, energy use, gender of the household head, mobility, use of tools and appliances, housing materials, and household size. The changes in outcomes will be captured by using baseline and follow-up surveys of households, businesses, and communities conducted, respectively, before and after the line extensions are completed.

Propensity score matching, a statistical method of matching based on multiple factors (Rosenbaum and Rubin 1983), was implemented in three stages. In the first stage, we applied nearest-neighbor matching with replacement and used existing census and global positioning system (GPS) data from the National Bureau of Statistics (NBS) as well as data from the Tanzania Electric Supply Company (TANESCO) to identify three potential comparison communities for each intervention community. NRECA International (NRECA), the firm contracted to carry out various baseline surveys for this evaluation, then implemented a community survey in the 182 selected intervention communities and 546 potential comparison communities. In the second stage of propensity score matching, we used data collected in the community survey and applied a matching without replacement method to identify one matched comparison community for each intervention community. NRECA then conducted a household survey in the 182 intervention communities and 182 comparison communities. Since the completion of the household surveys, we learned from MCA-T and TANESCO that four of the 182 surveyed intervention communities will no longer receive new lines under the T&D activity. Consequently, we are excluding these four communities from the evaluation, which brings the total number of intervention communities to 178. We used the results of the household survey for a third and final stage of matching of households in the intervention and comparison groups. A detailed technical discussion on this final stage of matching is provided in the appendix of this report.[6]

2. Random assignment evaluation design for the financing scheme initiative

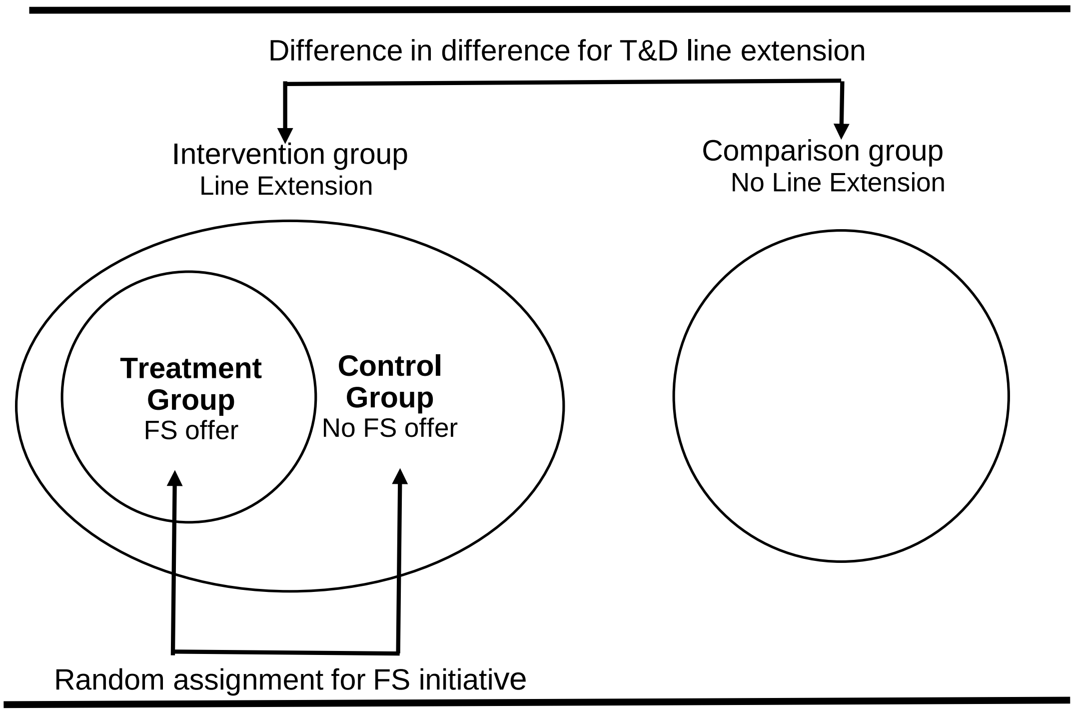
The FS initiative has been implemented only in the communities covered by the T&D line extensions. It will cover all six regions in the T&D evaluation as well as the Kigoma region, which was not in the original design and was added later. Therefore, the evaluation of the initiative is closely related to the evaluation of the T&D activity, as illustrated in Figure II.1. With the exception of the part of the FS sample in Kigoma, both the treatment and control groups for the FS initiative evaluation are selected from among the intervention communities for the T&D evaluation. In a public event on July 16, 2012, we randomly assigned 29 communities to the

not cover that region. The communities that received T&D lines in Kigoma will be included in the FS initiative evaluation using data from the follow-up survey.

[6] In Appendix Table C.2 of Chaplin et al. (2012), we present differences between the intervention and comparison groups means for over 200 variables. We find that the percentages of the differences that are statistically significant at various significance levels are consistent with what one would expect by chance alone. To help ensure that these remaining differences do not lead to biased estimates of impacts, we will use regression adjustment for other covariates. Key control variables will include the baseline measures of the outcome. This regression adjustment should also increase precision of the DID estimates by eliminating extraneous variation due to those covariates (see, for instance, Rubin 2007; Imai and Van Dyk 2004; Robins and Rotnitzky 2001; Rubin 1973).

treatment group that will receive the FS initiative.[7] This is out of a total of 192 communities- 178 in the six regions covered by the T&D evaluation and another 14 communities in Kigoma. The remaining 163 intervention communities constitute a control group that will not be offered low-cost connections. Mathematica hired a communication firm, Camco, to inform the 29 communities about the offer of low-cost connections and to encourage households to consider taking advantage of those offers.

Figure II.1. T&D evaluation: Overlap of line extensions and FS initiative



The design for the evaluation of the FS initiative has two implications for the evaluation of the T&D activity. First, when we estimate overall impacts of the T&D activity, we will also be capturing impacts of the low-cost connections as well as of the outreach work occurring through the communications campaign to reach the portion of our T&D intervention group that receives the FS initiative. Second, by excluding the FS treatment communities, we will be able to estimate impacts of the line extensions without the low-cost connections initiative when estimating the

[7] A total of 30 communities were randomly assigned to receive the FS initiative; however, two of the communities were in the Kigoma region, which was not covered by the T&D baseline surveys, but will be covered by the follow-up surveys. In addition, one community that was randomly assigned to receive the FS initiative during the public event will not be receiving new lines under the T&D activity (this is one of the four communities mentioned earlier that are no longer receiving the MCC funded lines). The decision to not provide new lines to this community, as well as to three other intervention communities, was made prior to random assignment. Consequently, these four communities are being dropped from the T&D evaluation. Any communities that changed status later will be included in the study with appropriate adjustments (Bloom 1984). No adjustments will be needed for intent-to-treat (ITT) estimates; however, we also plan to estimate treatment-on-the-treated (TOT) effects, which will adjust for crossover.

impacts of the T&D activity (although the resulting estimates will be somewhat less precise than our main results because of the smaller sample sizes).[8]

As noted above, we will also test to see if the more rigorous impact estimates are similar to the less rigorous ones estimated using simple pre-post comparisons. To do this, we will estimate impacts using pre-post methods for the T&D activity and the FS initiative and then compare those results to the estimates obtained using our more rigorous methods (described above). We will adjust our statistical tests of the differences to account for the correlations between the rigorous and nonrigorous estimates that result from their being estimated using the same data. This adjustment will improve our ability to assess whether the estimated differences are statistically significant.

C. Sampling

To provide data for the T&D evaluation (which covers FS), we conducted three baseline surveys prior to the implementation of the T&D activity: the Tanzania energy sector baseline community survey (or more succinctly, baseline community survey), the Tanzania energy sector baseline household survey (baseline household survey), and the Tanzania energy sector baseline enterprise survey (baseline enterprise survey). After conducting the baseline surveys, we realized that we could capture a broader array of enterprises from the household survey data than from the enterprise survey; for this reason, we dropped the enterprise survey.[9] In this section, we describe the sampling strategies applied for the community and household surveys.

1. Sampling for the baseline community survey

The baseline community survey was conducted in 182 intervention communities and 546 potential comparison communities in six regions. The primary sampling unit (PSU) for the community survey was a village (*kijiji*) in rural areas and a *mtaa* in urban areas.[10] The rural and

[8] In Chapter VI of Chaplin et al. [2012], we show equivalence between the intervention and comparison households at baseline after matching at the household level between the treatment and control groups. Since the control group households are a random subset of the intervention group, we would also expect to see baseline equivalence between households in the control group and those in the comparison group. This means that we can also use these data to estimate impacts of the new lines without FS.

[9] The baseline enterprise survey covered only 64 enterprises in 14 communities in just one region (Tanga). In contrast, the household survey, which also had a large number of questions about household-run enterprises, covered over 10,000 households in 364 communities in six regions. Although the enterprise survey was specifically focused on larger enterprises, we ended up getting more large enterprises in the household survey because of its larger total sample size and because most households had at least one enterprise. A key reason we found few enterprises in the enterprise survey is that we had planned to only survey enterprises that did not currently have grid electricity. However, we were not able to identify enough stand-alone enterprises that fit this description, so we ended up with a sample where about half had access to the grid. In addition, the largest enterprise found in the enterprise survey had only 6 employees. In contrast, the maximum number of paid employees in enterprises found in the household survey was 50, and 133 households had at least one enterprise with more than 6 employees.

[10] The Swahili word *kijiji* (plural *vijiji*) means village and refers to a rural administrative unit; *mtaa* (plural *mitaa*) translates to “street” and refers to the smallest urban administrative unit. Villages can be further subdivided into subvillages (*vitongoji*, singular *kitongoji*), which is the smallest rural administrative unit. Because the English word “street” could be confusing for a geographic area, throughout this report we use the Swahili words *mtaa* or *mitaa* to

urban communities covered by the community survey were selected in three steps. First, the evaluation team worked with MCA-T and TANESCO to finalize a list of communities (villages or *mitaa*) that were likely to receive new lines; we identified a total of 337 communities (Table II.2).[11] Second, we randomly selected 182 of those villages and *mitaa* to represent the intervention communities in the evaluation. This number was chosen to achieve the desired level of precision, as explained in our earlier design report (Chaplin et al. 2011). Third, as mentioned in Section B, we identified 546 potential comparison villages using propensity score matching and existing data. The potential comparison communities were chosen from among all of the non-intervention communities in the same region. Table II.2 presents the distribution of the intervention and potential comparison communities across the six regions in mainland Tanzania where the T&D activity is being implemented. The numbers of intervention and potential comparison communities sampled were chosen to have the same distribution across regions as the total population of intervention communities, as shown in Table II.2.

Table II.2. Number of intervention and potential comparison communities for the community survey, by region

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Intervention villages/*mitaa*				Non-intervention villages/*mitaa*		
Region	Total number of villages/ *mitaa*	Total number	Percent of total	Number sampled	Percent of total sampled	Total number	Number sampled for community survey	Percent of total sampled
Dodoma	658	73	22	39	21	585	117	21
Iringa	1,017	37	11	20	11	980	60	11
Mbeya	1,330	21	6	11	6	1,309	33	6
Morogoro	1,009	74	22	40	22	935	120	22
Mwanza	1,186	55	16	30	16	1,131	90	16
Tanga	1,269	77	23	42	23	1,192	126	23
Total	6,469	337	100	182	100	6,132	546	100

Note: The number of potential comparison communities in column 8 equals three times the number of intervention communities in column 5.

2. Sampling for the baseline household survey

The baseline household survey was conducted in 182 intervention communities and 182 matched comparison communities.[12] The 182 comparison communities were chosen from among the 546 potential comparison communities using propensity score matching based on the community survey data, as explained in the appendix. For the household survey, in urban areas we

refer to the urban communities in the evaluation. For the rural communities, we use “villages” and “subvillages” to refer to *vijiji* and *vitongoji*, respectively.

[11] The 337 villages and *mitaa* on our list were divided into 182 subprojects. Subprojects are units used by MCA-T and the implementing entities building the lines.

[12] During the household survey, we had to replace seven comparison communities because all households in those communities were within 30 meters of existing lines or were already connected and, thus, were not eligible for the survey.

continued to use a mtaa as the PSU. In rural areas, when a village had multiple subvillages, we used a subvillage (kitongoji) as the PSU; otherwise, we used the village as the PSU.[13, 14] In the intervention group, for a village with multiple subvillages, we selected the subvillage with the largest percentage of households expected to have access to the new T&D lines (as reported by the community leaders in the baseline community survey).[15] In each comparison village with multiple subvillages, we selected a subvillage that was matched to the population rank of the corresponding intervention subvillage. The purposive selection of the subvillage as the PSU in rural areas allowed us to achieve a much higher proportion of households in the sampling frame expected to have access to the new lines than we would have achieved had we used the village as the PSU. Without this purposive selection of subvillages, the evaluation would have needed a much larger sample of households to have reasonable confidence in detecting impacts. We did not need to identify a smaller PSU in urban areas because we expected that in urban areas receiving new lines, almost all households will have access.

Because we selected intervention communities that are expected to have a high percentage of households with access to the new lines, results from the evaluation will not generalize to households in communities where only a small fraction of households have access to electricity. However, focusing on communities with better access to new lines is better suited to inform future policy decisions about electrification, because future projects would build on the T&D activity and move closer to providing access to electricity for most or all households in the long term. Consequently, estimating impacts for communities where a greater percentage of households have access to electricity will be more policy relevant than estimating impacts for subvillages where only a small fraction of households have access.[16]

For the baseline household survey, in addition to identifying the communities, we had to sample households. For each intervention and comparison community (village, subvillage, and *mtaa*) selected for the baseline household survey, a list of all households residing in the community was created; this list also identified whether a household was already connected to the grid or near an existing line.[17] We do not expect that households that are already connected to the grid or close to an existing line will connect to the new lines. Consequently, we excluded them from the household survey sampling frame. The remaining households on the list constituted the sampling

[13] At the time of the baseline household survey, about 72 percent of the communities in the evaluation were classified as rural and the remaining 28 percent were urban (*mitaa*).

[14] This also mattered for the FS initiative, which was only supposed to be offered in one subvillage per village—the one selected for the household survey.

[15] Here, access to the electricity lines implies that the household is within a certain distance from the new low-voltage lines. Households or businesses within this distance are eligible for connection at a basic rate. Entities farther away must pay for additional poles. Currently, the distance is 30 meters.

[16] In estimating impacts of the T&D activity, we will use weights to adjust for sampling, nonresponse, and matching so that the estimated impacts represent impacts on household outcomes in communities where large fractions of households are receiving the new T&D lines (Angrist and Pischke 2009, p. 91–92; Pfefferman 1993).

[17] According to data we received from TANESCO, about one-third of the intervention communities where the new lines are being built already had existing lines. TANESCO provided us with these data to help us develop a sampling frame for communities in the intervention group.

frame for each community. In both the intervention and comparison communities, we sampled the same fraction of households from each PSU, which meant that we interviewed more households in the larger communities.[18] Within the intervention group communities, we oversampled households with a small house (these were being considered for a targeted subsidy pilot activity that was not implemented), as discussed in the appendix of this report. We will weight our estimates to adjust for this oversampling so that the estimates are representative of all households in these communities that were not already connected or within 30 meters of an existing line.

D. Baseline data collection

Using the sampling strategy described in the preceding section, three baseline surveys were conducted at the community, household, and enterprise level to support the evaluation of the T&D activity. MCA-T contracted with NRECA to administer all three surveys. NRECA developed the survey instruments, with input from MCC, MCA-T, and Mathematica. Table II.3 presents summary information on the respondents to each survey and the time period when each survey was in the field. The community survey was conducted first, over a seven-week period from April 18 to May 28, 2011. Data collection for the household and enterprise surveys started on August 15, 2011. The enterprise survey, a much smaller data collection effort, was completed within three weeks, on September 3, 2011. The household survey required a total of 14 weeks of field work and was completed on November 20, 2011.

There was a potentially important difference in how the intervention and comparison group household surveys were conducted. The data collection team prepared lists of all households in the sampled intervention and comparison communities; these lists were used to produce the sampling frame for the household survey. In the intervention communities, the list of households was prepared while the community survey was being fielded; for the comparison communities, it was prepared the day before the household survey was administered in a particular community.[19] This could have generated differences in results if large fractions of households moved during the months between the community and household surveys in the intervention group. However, as discussed in Chapter IV, our analysis suggests that a fairly small percentage of households moved

[18] In theory, we could have achieved more precise results had we randomly sampled communities proportional to their size and then sampled an equal number of households in each community (Lohr 1999). However, we lacked data on subvillage size when we drew our sample of villages and *mitaa* for the community survey. We could have sampled villages instead of subvillages, but the data on village size were dated, from the 2002 Census. In addition, based on the community survey data, we now estimate that only about 33 percent of the households in the target villages will have access to the new lines, compared to about 69 percent of the households in the subvillages we selected for the household survey. Thus, selecting subvillages with high access more than doubled the estimated fraction of households with access to the new lines in our intervention group village sample.

[19] The difference in the timing of the household listing in the intervention and comparison communities occurred for a number of reasons. The community and household surveys were conducted in the same intervention group communities. Consequently, for the intervention group, NRECA was able to carry out the household listing and the community survey at the same time. Moreover, we needed to identify households with small (no more than two rooms) versus large houses for the planned subsidy pilot activity in the intervention communities, so that we could oversample subsidy-eligible households. As a result, the listing of households in the intervention communities had to be carried out long before the fielding of the household survey. In contrast, for the comparison group, the community survey was conducted in three times as many communities as the household survey, and data from the community survey were used to select the communities where the household survey was administered. Consequently, it was not possible to do the household listing at the same time as the community survey in the comparison communities.

during this time in the intervention group. In addition, results discussed in Chapter VI of Chaplin et al. (2012) suggest that dropping these households does not affect the comparability of the intervention and comparison group households.

The baseline community survey was administered as planned and data were collected from all target communities. For the baseline enterprise survey, a total of 59 businesses responded to the survey (32 in the intervention group and 27 in the comparison group) compared to the target of 64 businesses.

Table II.3. Purpose, respondents, and timing of baseline surveys for the Tanzania energy sector evaluation

Survey	Purpose	Regions	Target sample size	Respondent	Start and end date
Baseline Community Survey	Collect community-level data at baseline; also used to identify matched comparison communities for the T&D evaluation	Dodoma, Iringa, Morogoro, Mbeya, Mwanza, Tanga	182 intervention, 546 comparison communities	Community leaders	April 18–May 28, 2011
Baseline Household Survey [a]	Collect baseline data on households for the T&D and subsidy pilot evaluations	Dodoma, Iringa, Mbeya, Morogoro, Mwanza, Tanga	11,648 households in 182 intervention and 182 comparison communities	Key female and male members of household	Aug 15–Nov 20, 2011
Baseline Enterprise Survey	Collect baseline data on small, medium, and large enterprises for the T&D evaluation	Tanga	64 enterprises in the intervention and comparison areas	Owner/ operator of the business	Aug 15–Sep 3, 2011

[a] All households in the sampled intervention communities were listed, and information on eligibility for a planned subsidy-pilot activity was completed, when the baseline community survey was administered in April–May 2011. This list was used to produce the household survey sampling frame for the intervention group.

In Table II.4, we present the distribution of the baseline household survey sample based on a data collection report prepared by NRECA, and in our final analysis sample by intervention status. NRECA's data collection team provided data on 10,298 households with complete survey data (4,767 in the intervention group and 5,531 in the comparison group). They reported an overall response rate for the household survey of 91 percent (86 percent for the intervention group and 99 percent for the comparison group). The regional distribution of the intervention and matched comparison communities where the baseline household survey was conducted is shown in a map in Figure II.2.

The final analysis file drops four of the intervention group communities where NRECA collected data and 88 of the intervention group households with completed surveys shown in Table II.4. As noted earlier, four communities were dropped from the intervention group because they will not be receiving new lines. These communities had 38 households. Another 41 households in the intervention group were dropped because they could not be matched to the household listing.[20] Six other households were dropped because they were duplicates. Thus, our final pre-matching sample size is 10,213 households from 178 intervention and 182 comparison communities.

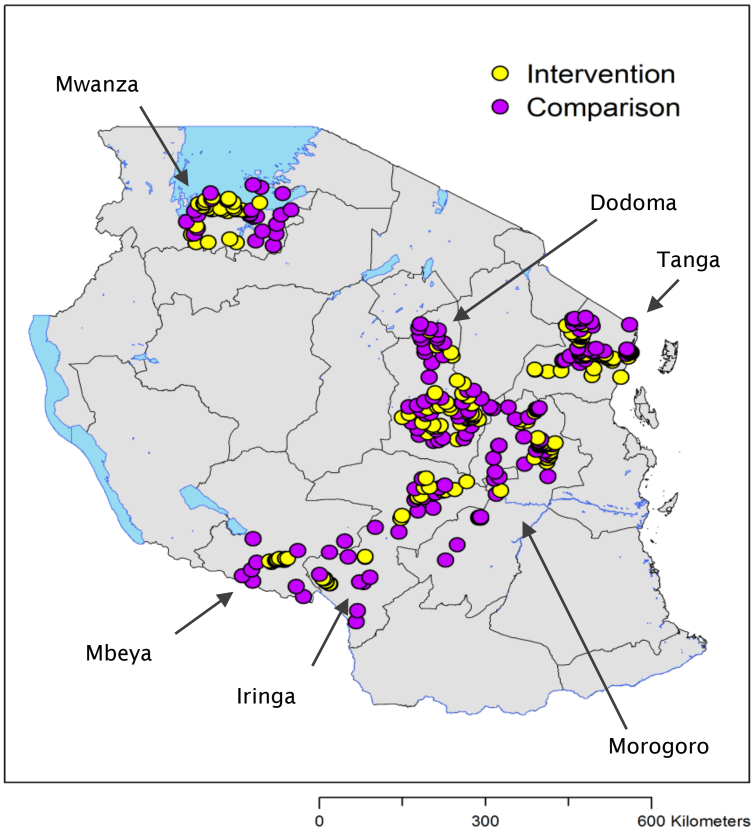
Table II.4. Baseline household survey: Matched intervention and comparison sample versus data from NRECA

	Intervention group		Comparison group	
Region	Total number of villages/*mitaa*	Number of households interviewed	Total number of villages/*mitaa*	Number of households interviewed
Data from NRECA	182	4,767	182	5,531
Not receiving new lines under the T&D activity	4	38	0	0
Could not be merged to household listing	0	41	0	0
Duplicate records	0	6	0	0
Not matched in propensity score analysis	0	3	0	0
Matched sample for T&D evaluation	**178**	**4,679**	**182**	**5,531**

Source: NRECA and Tanzania Energy Sector Baseline Household Survey

[20] We needed to merge the household survey data with the household listing in order to calculate their selection probabilities, which depended on eligibility for the subsidy pilot intervention as estimated during the household listing. Unfortunately, NRECA did not provide us with a numeric household ID to do this matching. Hence, we had to use the community ID and, within each community, the names of the heads of the households. These names often changed between the listing and the household survey, so we used approximate matching. Even with these approximate matches, we were unable to locate 41 households.

Figure II.2. Regional locations of the intervention and matched comparison communities in Tanzania



Sources: Tanzania Energy Sector Baseline Household Survey and Global Administrative Areas Database.

Notes: This figure is for descriptive purposes only, as it was not always possible to determine the accuracy of the GPS data. We have mapped 176 of the 182 intervention communities, and 181 of the 182 matched comparison communities in this figure. We plan to collect the GPS data again during the follow-up survey in a format that can be used for analytic purposes.

Descriptive statistics for intervention and comparison group households in our pre-match sample are presented in Appendix Table C.1 of Chaplin et al. (2012). We dropped three more intervention group observations after conducting propensity score matching at the household level because we could not find suitable matches in the comparison group (see the appendix of this report for details). Thus, our final baseline sample size post-matching was 10,210 households. Descriptive statistics for intervention and comparison group households in the post-match sample are presented in Appendix Table C.2 of Chaplin et al. (2012).

E. Statistical power

Even elegant study designs may be undermined by inadequate sample sizes. Large sample sizes protect against a "false negative" finding—that is, the failure to detect true program impacts simply because the study lacks statistical power. The sample sizes for the household survey are large and should provide sufficient power to detect household-level impacts of policy-relevant magnitude. The sample size for the community survey is relatively small, so results based on that survey will be more illustrative. In discussing the statistical power for estimating impacts, we focus on the household survey here.

In Table II.5, we present the minimum detectable impact (MDI) for the evaluation of the T&D line extension and FS initiative for a number of key outcomes. The MDI is the smallest true impact that can be detected with a given level of power. Thus, the smaller the MDI, the better it is for an evaluation. We used data from the baseline household survey to calculate the MDIs, accounting for clustering at the community level.[21] The MDIs are based on the assumption that the control variables to be included in the impact estimation model would explain half of the variation in the outcomes (that is, an R-squared of 0.5).[22] In addition, we calculated the MDIs with and without the weights that account for survey sampling, nonresponse, and matching at the household level. As shown in Table II.5, MDIs without the weights can be as much as 16 percent smaller, but most are less than 5 percent smaller. However, because applying the weights makes the sample representative of the underlying population, in discussing the MDIs below, we focus on weighted calculations.

Our ability to detect impacts varies across variables. For the outcomes presented in the table, the average MDI is 9 percent of the standard deviation of the outcome for the T&D line extension results and 18 percent for the FS initiative results.[23] We expect to be able to detect impacts as small as 15.4 kilograms on solid fuel use, as small as 0.07 hours on children's hours of study after sunset, as small as 4.8 percentage points on the $1-a-day poverty rate, and as small as 404,000 TZS on annual household consumption. For consumption, this is about 14.6 percent of the intervention group mean observed at baseline. For the other outcomes, these MDIs represent smaller fractions of the means.

It is important to recognize that the impact of the T&D line extensions probably hinges critically on the take-up rate which, in turn, depends on the percentage of households in the intervention communities that have access to the new lines, as well as on the percentage of households with access that actually connect to the lines. If either of these percentages is low, expected impacts of being in a community with access to the lines are also likely to be low. For example, suppose that using electricity increases consumption by 15 percent, or by about 415,000 TZS. If only 70 percent of the intervention group households have access to electricity, and only half of the households with access install a connection, then on average we would expect to see an increase of only 145,000 TZS in consumption in the communities that get access, compared to those that do not. This is much less than the MDI for consumption reported in Table II.5.

---

[21] Clustering occurs because residents of the same community are likely to face similar, unobserved (by the researcher) random shocks that affect the outcomes. This results in greater correlation of outcomes among households in the same community than can be explained by observed variables that will be included in the impact estimation model.

[22] If the R-squared statistic drops by half, to 0.25, then the estimated MDIs increase by about 22 percent (compared to those in Table II.5). If R-squared goes to zero, then the estimated MDIs increase by about 41 percent.

[23] These MDIs in standard deviation unit are very close to the clustering-adjusted MDIs presented in the evaluation design report (Chaplin et al. 2011).

Table II.5. Minimum detectable impacts (MDI) for T&D evaluation: Line extension and FS initiative

Variable	Intervention group		Minimum detectable impacts: T&D line extension		Minimum detectable impacts: FS initiative	
	Mean	Std dev	Weighted	Unweighted	Weighted	Unweighted
Household-level outcomes						
Monthly amount of solid fuel used (kg)	151	209	15.42	15.17	40.11	36.77
Average hours/day spent studying after sunset, members ages 5–24 (hours)	0.66	0.87	0.073	0.070	0.114	0.115
Adult has had health problems in past 7 days (%)	45.2	49.8	5.22	5.07	11.56	11.58
Child died if any born alive in last two years	8.6	27.9	2.71	2.61	5.66	5.72
Number of tools and appliances	7.0	4.7	0.46	0.39	0.65	0.63
Household has no IGA (%)	29.5	45.6	4.35	4.22	10.00	9.52
Average number of female-operated income-generating activities (IGAs) if household has IGAs	0.47	0.58	0.045	0.041	0.080	0.076
Average number of male-operated IGAs if household has IGAs	0.63	0.71	0.061	0.058	0.099	0.096
Makes less than $1 US income per capita per day (percent)	71.7	45.0	4.83	4.51	9.78	9.22
Annual household consumption (TZS)	2,769,502	3,882,798	403,656	352,637	710,964	664,299
Annual household consumption (USD)	1,756	2,462	256	224	451	421

Source: Authors' calculation based on Tanzania energy sector baseline household survey data.

Note: The analysis accounts for clustering by community. To calculate the MDIs, we assumed a confidence level of 95 percent, two-tailed tests, 80 percent power, and R-squared = 0.50. The sample sizes for each outcome shown in the table are available by intervention and treatment status in Appendix Tables C.2 and C.3 of Chaplin et al. (2012), respectively. The MDIs for T&D line extension are calculated using the matching weight (W^M_i) described in Section 2 of the appendix of this report. The MDIs for FS initiative are calculated using the pre-match weight (W_i), described in Section 1 of the appendix of this report.

The MDIs are larger for the evaluation of the FS initiative because the household sample size (including both the treatment and control groups) is reduced by half compared to the overall sample size for the T&D evaluation (intervention and comparison groups combined). Also, although the intervention group constitutes about half of the T&D sample, the treatment group constitutes only about 15 percent of the FS sample. This lack of balance also increases the MDIs. For example, for solid fuel use, we will be able to detect an impact of at least 40.1 kilograms; for children's study hours after sunset we will be able to detect an impact of at least 0.11 hours, and for the $1-a-day poverty rate, we should be able to detect an impact of at least 9.8 percentage points. These MDIs reflect the assignment of 27 non-Kigoma communities to the treatment group for the FS initiative and 149 to the control group. Adding in the Kigoma region reduces the MDIs slightly by adding 2 treatment and 12 control communities.[24]

For the FS initiative, the expected impacts of being connected to the electricity grid depend in part on the fraction of the control group that gets connected to the grid. We expect that a similar fraction of households in the treatment and control groups will have access to the new lines. However, only the treatment group was offered the low-cost connections. We estimate that there are about 6,340 households in the treatment group that are potentially eligible to receive the low-cost connection (that is, not connected to an existing line).[25] MCC made about 5,800 low-cost connections available to these communities. About one third of these connections were used.[26]

The calculations above are based on the assumption that our response rate at follow-up is similar to our response rate at baseline. This is possible since we will be including the baseline nonrespondents in our follow-up survey sampling frame, and in our final analyses if they do respond.[27] Nevertheless, the response rates at follow-up may be significantly lower than at baseline, because of mobility. We will try to survey households that migrate, but it may be difficult to locate some of them.[28] Consequently, as Figure II.3 below shows, the MDIs will change very little as a function of the response rate as long as the response rate is above 70 percent and we do not lose any communities from the sample.[29]

[24] The reduction in the MDIs will be limited to some extent by the fact that we will lack baseline data in those communities.

[25] The estimated total number of households includes 140 households in the two communities in the Kigoma region that are not part of the T&D evaluation.

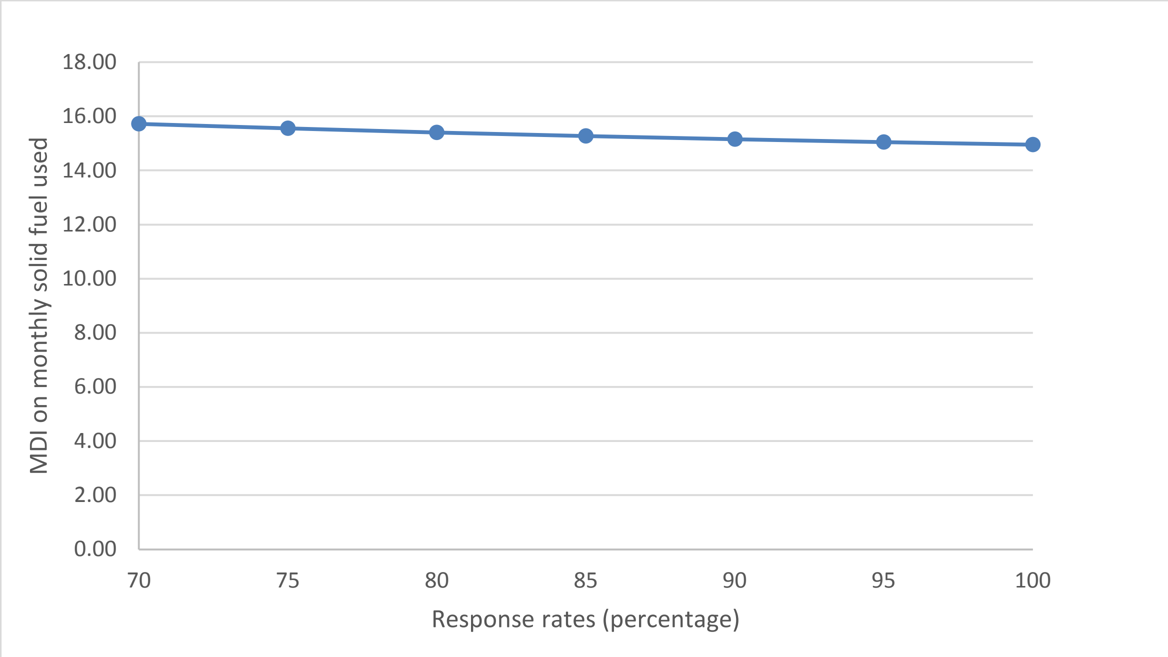
[26] We had originally planned a second round of random assignment to use the additional low-cost connections, but this plan had to be cancelled due to a lack of time before the end of the compact.

[27] We will lack baseline controls for these households, but can still include community-level averages as control variables for them.

[28] The survey firm EDI was selected in part because of their expertise in following migrant households in Tanzania. Indeed, they performed a study following households over 12 years. In comparison, our follow-up survey is only 4 years after the baseline.

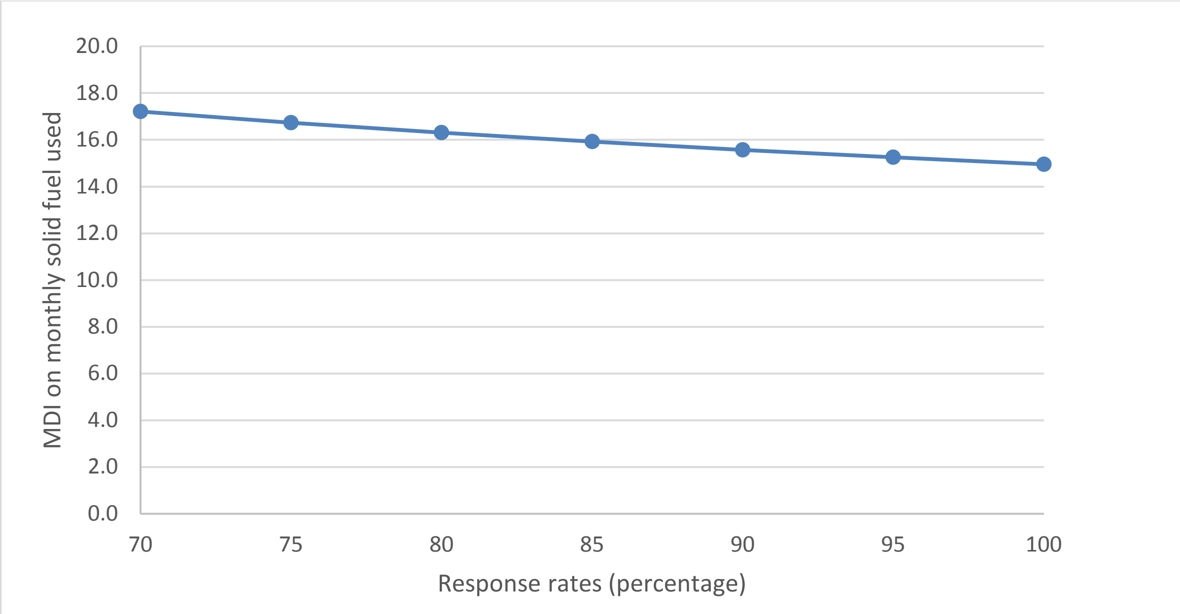
[29] These calculations were done by estimating an interclass correlation (ICC) using the results from Table II.5 and assuming an even number of households per community. The ICC was 0.0883. We then used that ICC to simulate the MDIs shown in Figure II.3.

Figure II.3. Minimum detectable impacts (MDI) of the T&D activity on monthly solid fuel used by household-level survey response rate



Even if a few sampled communities are lost, change in the MDIs should be moderate. Figure II.4 shows estimates of what would happen if up to 30 percent of the communities do not respond to the follow-up survey (which seems very unlikely).

Figure II.4. Minimum detectable impacts (MDI) of the T&D activity on monthly solid fuel used by community-level survey response rate



The evaluation's ability to detect impacts on outcomes based on the community survey will be weaker because of smaller sample sizes. Data from that survey will be used for case studies and are expected to provide illustrative findings that will inform other components of the evaluation. As such, this lack of statistical power should not be critical.

Since we will be estimating impacts on a large number of outcomes, we must be mindful of the statistical problem of "multiple comparisons." This problem may arise when researchers estimate impacts on a large number of outcomes: at least a few of the estimates are likely to be statistically significant by chance, even if no true impacts occurred. We will take a balanced approach to addressing the multiple comparisons problem (Schochet 2009). This will entail a tradeoff between reducing the likelihood of getting "false positives" (that is, finding statistically significant impacts by chance even when no true impacts exist) and maintaining our ability to avoid "false negatives" (that is, the statistical power to avoid incorrectly inferring no impacts when true impacts exist). First, we will estimate impacts on rate of connection to the electricity grid. If there is no impact on connection rates, it would be reasonable to assume that impacts on any other outcomes are unlikely to be caused by the new lines. Second, for the remaining outcomes we will work with MCC to pre-specify a parsimonious set of outcome domains, and specify one or a few primary outcomes in each domain. Table II.6 below constitute the starting point for this process. The primary outcomes will be the basis for tests of the main hypotheses. By limiting the number of main hypotheses being tested, this approach will reduce the likelihood of finding impacts by chance alone, without significantly undermining the evaluation's statistical power to detect true impacts.

We also plan to estimate impacts on a large number of additional supplementary outcomes in each domain. Some examples of these are also shown in Table II.6. Examples of additional supplementary outcomes we may include in our final report can be found in Appendix C of our baseline report (Chaplin et al, 2012). Impacts on these and many other supplementary outcomes will be presented in our final report but will need to be interpreted with more caution. We will only highlight the findings for the supplementary outcomes if we find statistically significant impacts on the primary outcomes or if we find a credible pattern of statistically significant impacts on the supplementary outcomes.

F. Qualitative analysis

This section describes the design of the qualitative evaluation of the T&D activity and the FS initiative, including the study domains, research questions, data sources, and analytic approach.

1. Qualitative study domains and research questions

The qualitative evaluation team selected a number of research questions in collaboration with MCC and MCA-T. The evaluation was designed to produce insights into the following questions, using data from interviews with key informants at the national, regional, and community levels and from focus group discussions with households. The principal study domains were project implementation (successes, challenges, and sustainability) and stakeholder outcomes (access to electricity, energy costs and benefits, health and education, productivity, and gender.) (Details of the questions in each domain are shown in Appendix A of Miller et al. [2015].)

Table II.6. Primary and secondary outcomes by domain for household-level impact analysis

Domain	Outcomes
Connection rates	
Primary outcomes	Overall connection rate to grid electricity
Examples of secondary outcomes	Connection rate to MCC lines; connection rate to non-MCC lines built since 2011; connection rate to non-MCC lines built before 2011
Composition and mobility	
Primary outcomes	Change in community size
Examples of secondary outcomes	Out-migration rate from community; number of new migrants to community; number of new households created, changes in outcomes for migrants
Energy use	
Primary outcomes	Electricity use; liquid fuel use
Examples of secondary outcomes	Solid fuel use; tools and appliance use; consumption of amount of light (in lumen-hours); cost per unit of light; total cost of light consumed; local pollution (soot)
Education and child time use	
Primary outcomes	Children hours studying at night
Examples of secondary outcomes	Enrollment in school; hours doing chores
Health and safety	
Primary outcomes	Adult sick in last 7 days; youth sick in last 7 days; child sick in last 7 days
Examples of secondary outcomes	Availability of vaccines; distance to closest health center; perceived safety at night
Business and adult time use	
Primary outcomes	Paid employment; number of income-generating-activities (IGAs)
Examples of secondary outcomes	Hours worked; types of IGAs; IGA open at night
Economic well-being	
Primary outcomes	Household non-electricity consumption
Examples of secondary outcomes	Household income; household assets; consumption by type; property values; poverty
Gender differences in impacts	
Primary outcomes	Perceived safety at night; hours of household chores
Examples of secondary outcomes	Paid employment; business open at night

Questions related to project implementation:

- **Implementation**. How were the T&D activity and FS initiative implemented?
- **Successes and challenges.** What are the implementation successes and challenges?
- **Sustainability.** How sustainable are the implemented components of the energy project perceived to be?

Questions related to stakeholder outcomes:

- **Connecting.** Why did households, businesses, and community institutions decide to connect or not to connect to the newly available electricity lines and to take advantage of the financing scheme?
- **Outcomes.** What are the potential impacts of increased access to electricity on economic, education, health, safety, migration, and related outcomes? What are the mechanisms by which these changes might occur?
- **Gender.** Do the changes in outcomes from access to electricity differ by gender? Why do these differences occur?

2. Data

We collected qualitative data about implementation successes and failures as well as perceptions about how implementation affected outcomes of households, businesses, and other community entities. We collected qualitative data one time, a few months after the T&D activity and FS initiative had been fully implemented. The timing allowed key stakeholders and respondents to reflect upon the implementation process as well as upon changes that had occurred since the project was implemented and customers gained access to grid electricity. The data were collected in eight communities, all of which received new T&D lines. Half had been offered FS and half had not.

a. Data sources

We used a variety of qualitative data collection methods and data sources to fully explore our research questions. Table II.7 describes the respondents, the method of data collection (key informant interview, focus group discussion, observation, or review of reports), and research questions addressed in the qualitative study.

Table II.7. Respondents, data collection methods, and research questions

Respondent type	Data collection method	Research questions
MCA-T and TANESCO staff • M&E and energy sector program directors at MCA-T • MCA-T liaison at the TANESCO headquarters and other TANESCO staff at the district level	7 Interviews (5 of TANESCO staff and 2 of MCA-T staff)	• Implementation of the project components • Informants' perceptions of the program's successes and challenges
Household members	Separate focus groups for men and women in each community. Household FGDs included both connected and not connected households. Focus groups included 74 men (36 connected and 38 not connected) and 78 women (39 connected and 39 not connected) in the 16 FGDs in the 8 study communities.	• Interest and ability of households to connect to lines; obstacles in getting connected • Perceptions of costs and benefits of accessing electricity • Experiences with and outcomes related to the financing scheme (for communities where it was offered) • How connected households use electricity and why they use it for some purposes but not for others • Women's particular points of view, and differential effects of the interventions on them
• Owners/managers of an income-generating activity (IGA) • Both formal and informal IGAs inside or outside the home • May or may not have paid employees • IGA owners are a subset of household members • They include male and female IGA owners, both connected and not connected to electricity	40 interviews (5 per community)	• Specific effects of the interventions on the IGA • IGA owners' perceptions of the value of electricity to the productivity and profitability of their businesses • Obstacles to connecting • The costs and benefits of electricity related to income generation • Different effects for male and female IGA owners
• Community leaders (village/mtaa head), school headmasters, and health facility directors (or other appropriate staff) who are positioned to describe their experience with electricity connectivity in their role • We included both connected and unconnected schools and health facilities, as well as any schools that were located in communities that were offered the FS initiative	24 Interviews (1 community leader, one school staff person, and one health facility staff person per community)	• Perceptions of the effects of electricity at the community level • The uses and benefits of electricity for schools and health facilities • Perceived effects on education and health outcomes • Quality of lines, power outages. • Other community-level outcomes such as economic activity and other externalities

Respondent type	Data collection method	Research questions
Not applicable	Detailed observations of community infrastructure by data collection team using an observation tool to record salient features of infrastructure	• Were schools and health centers electrified? • Was there evidence of illegal connections? • Were there local markets? How large, and were they electrified? • Have new businesses emerged? • What was the location of power lines relative to households and businesses?
Not applicable	Review of implementation reports and monitoring and evaluation data on implementation and outcomes by researchers	• This review of documents and data provided by the implementers and MCC supplemented primary qualitative data to answer the research questions and provided some "lessons learned" regarding implementation.

b. Site selection

We collected qualitative data in eight T&D project communities—two from each of four regions: Mwanza, Dodoma, Iringa, and Tanga.[30] These regions were selected purposively by MCC and Mathematica to include a range of different conditions for the qualitative study. Specifically, Mwanza, located on Lake Victoria, has a social and economic profile quite different from other regions in Tanzania, and Tanga is the most industrialized region of the country. Dodoma was suggested by MCC because the engineering model applied to design the line extension (by Pike Electric) was different in Dodoma than in the other regions. Finally, we selected Iringa as the fourth region because the baseline household survey suggested that communities in Iringa averaged more households than communities in other regions, providing a larger pool of potential qualitative study respondents from which to choose. One urban and one rural community were randomly selected from each region.

Data collection took place in those intervention communities where MCC-funded lines had been built. We purposively selected communities to vary on key characteristics, including rural/urban status and those offered and not offered the FS. We also selected communities with businesses and communities that had at least some connected and some not connected households.

c. Development of data collection tools, and data collector training

We developed key informant interview guides, focus group protocols and observation tools in English, guided by the study domains, research questions, and program logic model identified as being of key interest in the evaluation. MCC and MCA-T reviewed the data collection tools and provided feedback, and we revised the tools accordingly. CSR Group Africa (CSR), our local qualitative research partner, contracted professional translators to translate the data collection tools

[30] To maintain respondent confidentiality, we have omitted names of the specific communities where we collected data. We do, however, present community background characteristics in Chapter IV to provide context for the outcomes discussed in that chapter.

into Kiswahili. The translated tools were reviewed during the in-country training, and also by Mathematica researchers who speak Swahili.

Together with CSR staff, we conducted a three-day training workshop in Dar es Salaam for the data collection researchers. The training developed a common understanding of the instruments, the data collection protocols, and the supporting systems for field operations. We led sessions explaining the MCC energy project activities, the scope of the full evaluation, and the sampling and recruitment procedures for the qualitative activities, and provided examples of high quality transcripts of focus group discussion (FGDs). The team completed detailed reviews of the qualitative protocols and guides, conducted mock interviews/FGDs, and reviewed the Kiswahili translations of the guides and protocols.

Once trained, the group pre-tested the guides and field procedures in an urban community in the Morogoro region. The teams made courtesy calls to the regional and district offices to obtain the necessary permissions and then met with the community leader in each site, who helped make arrangements for the focus groups and identified a nearby school for the school director interview. The team completed the community leader and school director interviews, screened and recruited households and owners/managers of businesses or IGAs for focus groups, and held three focus groups. After the pre-test, the data collection guides and study procedures were revised; for example, we replaced focus groups for business owners with interviews because business owners were reluctant to leave their places of business during the day.

d. Data collection procedures

The field staff divided into two teams, each covering four communities. In each location, the field staff obtained the necessary approvals before proceeding to the community leader. The team asked the community leader for help in conducting observations, identifying the location of schools and health facilities, and identifying IGA respondents. The field teams also walked through the local markets to identify and recruit IGA respondents. Household focus group participants were randomly selected from one of two lists. The field team first asked the community leader to use the community register of households. If that list was not available (which happened in two communities) then the field team used the list of households provided by Mathematica from the baseline household survey conducted in 2011. The team was not able to find all households from the Mathematica list because some households had migrated, and also because many individuals change which of their names they use when reporting on surveys.[31] Consequently, when the two lists were unavailable or ineffective, the team developed a new list of households by walking through the community. When the team had a list of potential participants for the household focus groups, it randomly selected the households to survey.

In each community selected, we conducted the community observation and also interviewed the highest ranking elected or appointed community leader for the subvillage or village, or his/her deputy if the leader was not available. We interviewed the school director and the health center

[31] Individuals often have a large number of names.

staff in each community.[32] We also conducted two focus groups (one for male household members, the other for female) and five interviews with IGA owners in each community. Interviews with MCA-T staff were conducted at the national level only. TANESCO staff interviews included one national-level interview and four interviews with district-level staff, one in each of the four study regions.

The field team completed the community observation tool by interviewing the community leader as well as walking through the community to assess characteristics of the location, such as the total number of businesses, the number of connected businesses, and the condition of the power lines. If there was any conflict between the community leader's answers and what the field team observed, the field staff noted it in the observation tool. We also reviewed implementation reports for the T&D activity and the FS initiative and used monitoring and evaluation data from stakeholders.[33]

Once participants were recruited, the interviewer sought consent to collect data and digitally record the interview. After the interviews, the field staff transcribed all digital recordings into Kiswahili transcripts, which were then translated into English. Finally, CSR staff checked all transcripts with the audio recordings to ensure quality and accuracy before sending the transcripts and audio files to Mathematica.

3. Analytic approach

We analyzed the qualitative data by first reading and rereading all transcripts. Next, we developed analytic codes and themes. We created a separate database for each type of transcript (male FGDs, female FGDs, health facilities, schools, community leaders, TANESCO, observation forms, and MCA-T). We asked CSR to check the audio files and translation if we needed more clarity to understand the transcripts. Next, we coded transcripts line by line using NVivo 10. Once coded, we compared the themes and codes by respondent type as well as by region and location. The comparison of data allowed us to identify differing responses and, in some cases, the reason for the differences, as well as to validate findings based on repeated reports across respondent types and locations. Results from our qualitative analyses are presented in Miller et al. (2015).

G. Interim impact analysis

As discussed above, we had planned to do an interim impacts analysis based on data on community-level connections gathered by TANESCO over a number of months. We finished a draft of the report but were later told that the total number of connections was much lower than the community-level data from TANESCO implied. We worked with MCA-T and TANESCO to do additional checks that confirmed that the original data overestimated total connections by at least 60 and perhaps over 100 percent. Based on this information, MCC decided to not release the

[32] Only three of the communities we selected for the household focus groups had health facilities. Two were in FS communities and only one of the two was connected to the grid. In order to obtain a total of eight interviews with health facilities, we interviewed five more in other communities that also got new MCC-funded lines: four were connected to MCC-funded lines, one relied on solar power, and three used several energy sources on a piecemeal basis.

[33] These reports are not publicly available so are not referenced here.

report. Details of the problems with those data are provided in Chaplin and Mamun (2015). In this section, we describe our original plans for using those data.

1. Outcome variables

We obtained community-level administrative data from TANESCO on the number of connections to the national electric grid in the 178 intervention communities and the 182 comparison communities described above.[34] The data on number of connections is at the community level and contains information on whether the connections are to MCC-funded lines or other (non-MCC-funded) lines. We used connections data from four available data points—January, June, and December 2013, and June 2014. This enabled us to look at the latest data available on connections (from June 2014) as well as at changes over time. Note that in January 2013, the MCC-funded lines may have been energized in only one of the seven regions (Dodoma) covered by the T&D activity, and the newly built lines in the other regions were energized over time during the period covered by our data.

Using the data from TANESCO, we defined two types of outcomes: (1) whether the community is connected at all (recorded as one if at least one household in the community is connected, zero otherwise),[35] and (2) number of connections in a community. For both of these types of outcomes, we examined three different measures of connections: connections to MCC-funded lines, connections to other (non-MCC) lines, and (3) their sum, connections to any lines (MCC funded or other).[36]

We estimated impacts on connections to the new MCC lines as well as to existing non-MCC lines. We examined connections to non-MCC lines because about a third of the communities in the study had existing non-MCC lines covering parts of these communities before the MCC lines were built, and because the FS initiative was available to all households in each treatment community—not just those with access to the new lines. Thus, the introduction of the new lines and the FS initiative could impact connections to existing lines in at least four ways. First, the introduction of new lines could reduce connections to existing non-MCC lines.[37] Second, people

[34] As noted earlier, we also received community-level connections data from TANESCO for 14 communities in Kigoma. These communities were added to the T&D evaluation after the baseline survey had already been conducted. All 14 are in the T&D intervention group, and two of the 14 are in the FS treatment group. Even though TANESCO has collected connection data for these communities, they are excluded from the main analysis because we lack baseline household survey data for them. We included these 14 communities in some robustness checks.

[35] We used OLS estimation for this binary outcome, which implies a linear probability model. We used robust standard errors to account for heteroskedasticity, and adjusted for clustering by district.

[36] An alternative outcome might have been the ratio of the number of connections to the total number of households in the community. However, we only have the total number of households in each community collected during the baseline survey in 2011. We chose not to use the ratios as outcomes out of a concern that households may have migrated into or out of many of these communities. If that had happened, the denominator for calculating the ratio would be incorrect. We relied instead on regression adjustment to account for the number of households residing each community at baseline.

[37] The introduction of MCC lines could reduce connections to non-MCC lines in two ways. First, non-connected households may choose an MCC line over a non-MCC line. Second, some households may switch from non-MCC to MCC lines. Households that have premises close to both sets of lines could do this most easily. Others could move to

near to the existing lines may be influenced by those connecting to the new lines. Third, TANESCO may be more likely to serve customers of existing lines in communities with new lines because of economies of scale.[38] Finally, the FS initiative was available to households with access to the new lines as well as those with access to the existing non-MCC lines.

We also looked at impacts on connections over time between January 2013 and June 2014. These impacts are likely to increase over time during this period because in January 2013, when our data start, only a fraction of communities were energized. Impacts on connections due to the FS initiative may be delayed further compared to the T&D impacts since the FS initiative was implemented after the newly built lines were energized in the selected communities. Impacts of both the T&D activity and the FS initiative could increase over time during the period covered by our data if households needed more time to save money for the connection fee (for T&D) and wiring (for both T&D and FS). For T&D, the impacts could also decrease over time if new non-MCC lines were built in the comparison communities.

2. Control variables

To generate the variables we control for in the regression-adjusted impact analysis, we merged the community-level connection data with several baseline survey data sets.

- **Baseline community survey.** The baseline community survey was used to identify the rural/urban status of each community.
- **Baseline NRECA household listing**. To obtain a sample for the household survey, NRECA conducted a census and prepared a list of all households in each of the study communities. These household listings include information on which households were already connected to the electric grid at baseline or were within 30 meters of the existing electric lines. Such households are unlikely to connect to the new lines. We created two variables from these data. One variable had the total number of households in each community. The other variable had the number of households that were neither connected nor within 30 meters of an existing line. Because the intervention and the comparison groups differed at baseline in terms of community size, these variables could help us account for the difference and reduce potential bias. In addition, because of the large amount of variation in community size within the intervention and comparison groups, these variables could also help to improve the precision of our estimates.
- **Baseline household survey**. Although the T&D intervention and comparison communities appeared to be equivalent on many community characteristics available from the baseline community survey data, the groups differed on a number of household characteristics available from the baseline household survey data. These characteristics included annual household income, household assets, annual household consumption, percentage of

access the new lines. Our data do not allow us to identify which households might have connected in the absence of the new lines.

[38] This is a possibility, because TANESCO has limited resources for making connections in terms of both available staff and vehicles and, consequently, may want to serve a community only after a certain minimum number of connection applications have been submitted. The existence of new lines increases the likelihood that a community reaches this threshold.

households using grid electricity, number of appliances owned by households, and hours per month household spent watching TV. When analyzing household-level data in the baseline report for the T&D evaluation (Chaplin et al. 2012), we adjusted for these differences by reweighting the data on household characteristics using propensity score matching weights at the household level. But because we only used community-level data in the interim impacts analysis, we were not able to use the household-level weights in that analysis. Instead, for the interim analysis, we controlled for the six key household characteristics at baseline mentioned above, that had statistically significant differences between the intervention and comparison communities. We included these variables as controls because they are likely to be correlated with a household's decision to get connected to the electricity grid. We created these variables from the baseline household survey, averaging up to the community level, using weights that adjusted for nonresponse and for the fact that poor households were oversampled in the intervention communities. We also used these variables as controls in the regression-adjusted impact analysis of the FS initiative. As noted earlier, after learning about problems with the data, MCC concluded that they were not reliable enough to justify releasing the draft report based on them. Details about the data problems we found are presented in Chaplin and Mamun (2015).

This page has been left blank for double-sided copying.

III. KIGOMA SOLAR EVALUATION PLAN

In this chapter, we describe our plans for evaluating the Kigoma solar activity. The Kigoma solar activity includes many components. In particular, it covers provision of solar energy systems for certain public institutions and village markets, provision of resources for solar-powered night fishing systems for fishermen, and the sale of solar systems for household and small business use with financing through local credit institutions. Supporting activities include marketing of the solar systems and information on their benefits, training of installers, vendors, and end users, and maintenance and post-sale services. The Kigoma solar activity is designed to improve electricity coverage and consumption of electricity, which should in turn increase economic activities by businesses and individuals, as well as access to education and medical services, all of which will contribute to poverty reduction and economic growth.

A. Study design

A timeline for the Kigoma solar activity and evaluation is presented in Figure VI.1 in Chapter VI of this report. MCC hired an independent contractor to conduct the evaluation of the Kigoma solar activity. That contractor collected a first round of survey data in 2013. Mathematica was hired to collect and analyze a second round of data; in turn, we hired Economic Development Initiatives (EDI) to collect the data. This second round will be collected in 2015.

We propose to do a pre-post analysis looking at how key outcomes change over time for beneficiaries as the program is implemented between the first and second rounds of data collection, and at how those changes compare to changes for a comparison group that was not targeted by the program. The pre-post analysis and the contrast with the comparison group will not be rigorous for two reasons. First, for most of the beneficiaries, the pre-post analysis will rely on retrospective data instead of baseline data collected prior to the offer of the solar power services. Second, the comparison group was not chosen to be similar to the beneficiary group. Nevertheless, the results should provide suggestive evidence regarding the relationship between Kigoma solar activity and beneficiary outcomes.

The data will not include true baseline information for most of the program beneficiaries because the intervention had reached many beneficiaries before the first round of evaluation data were collected. More precisely, the program was supposed to be implemented between March 2012 and May 2013 but the first round of data were not collected until July 2013 and about two thirds of the "installations" were complete at that time (Busalama 2013). Respondents were asked for their recollection of conditions 12 months prior to the survey, but even that information does not represent a true baseline for all of the programs since it refers to July 2012, a few months after implementation was supposed to have started.

The data do not provide a true baseline for all components of the program, but they do provide baseline information for one component, the hospital refrigerators. The first round of data were collected after about two-thirds of the program had been implemented (Busalama 2013) but before the hospital refrigerators had been distributed. Hence, we have baseline information for that component of the activity.

Our evaluation will cover six types of direct beneficiaries of this activity as well as comparison group respondents for each of these types. The beneficiary types are (1) schools, (2) health facilities, (3) businesses in village markets, (4) fisherman, (5) businesses that received loans from local credit institutions to purchase solar systems, and (6) households that received loans from local credit institutions for solar systems. Round one data were collected from a total of 82 direct beneficiaries and a comparison group of about 40 entities of the same types. We plan to collect the second round of data from this same set of respondents in fall 2015, slightly over two years after the round one data were collected. Sample sizes by beneficiary type are provided in Table III.1 below.

Table III.1. Sample sizes for Kigoma solar evaluation

Beneficiary type	Total	Intervention	Comparison	% Intervention	% Comparison
School	15	10	5	12%	13%
SACCO Household	24	16	8	20%	20%
Market Business	20	12	8	15%	20%
SACCO Business	24	16	8	20%	20%
Dispensary	18	14	4	17%	10%
Health Center	9	6	3	7%	8%
Fisherman	12	8	4	10%	10%

The comparison group will not enable us to estimate impacts rigorously but will be helpful for assessing potential impacts of the activity. The comparison group respondents were identified "purposefully to ensure that they are located far from target villages and facilities of the program" (Busalama, 2013) though still within the Kigoma region and of the same entity types (school, household, business, and so on) as the intervention group.

Our ability to conduct this evaluation of the Kigoma solar activity will depend both on the quality of the round one data and on our ability to follow up with those respondents. We have shared the round one data with the data collection firm we are working with for the round two surveys. They have agreed to try to track all round one respondents who remain within Kigoma or migrated to one of the other study regions.

B. Data collection

Because we hope to compare outcomes from the round two data with outcomes from the round one data, we plan to ensure that the survey instruments used for round two data collection are as similar as possible to the instruments used in the first round. In addition, we propose adding questions on the following topics in the round two survey instruments:

- **Additional questions about energy use**. In the first round, respondents were asked about the levels of use of various sources of energy in the 12 months prior to the survey and at the time of the survey. We propose to add additional questions about reasons for use and non-use of these sources of energy, especially solar power.

- **Additional questions about health and safety.** We propose adding questions to the household survey about the perceived safety of the energy sources they reported using, and any perceived health effects that may result from the use of these sources.

Our survey instrument will have two sections—one that contains questions asked to all beneficiaries regarding use of solar power and alternative forms of energy, and another set of questions that varies by beneficiary type. The second set of questions will include items that are more targeted for a specific group of beneficiaries. For example, fisherman might be asked about how many fish they catch, whereas hospitals might be asked about what types of drugs they store in refrigerators. This approach would allow us to obtain the largest possible sample size for those questions that can be asked in the same way for all beneficiaries, while obtaining additional information that is specific to a type of beneficiary but still of value.

This page has been left blank for double-sided copying.

IV. T&D MIGRATION OF HOUSEHOLDS

In this chapter, we describe our plans to address migration issues in the T&D evaluation, which covers both the T&D activity and the FS initiative. Our baseline survey will enable us to rigorously measure impacts for households that were in the survey communities at baseline. This includes households that migrate out, as we will be able to follow them to their new communities. However, the baseline survey sample will not allow us to estimate impacts for in-migrants to those communities, since we will not have a good comparison for those households. Consequently, we are planning to conduct various analyses to help capture the potential for benefits for in-migrants.

During early planning visits for our evaluation and during the qualitative data collection period, we noticed a number of communities where the MCC lines households can connect to were being built or had been built in areas with very few homes. This could happen if developers purchase large plots of land that they intend to sell off as small parcels, each of which has a home, or if it is difficult to obtain property rights to build lines close to existing homes. In either case, it would not be surprising if many households end up migrating to get close to the new lines after they are built. If in-migration rates are low and/or similar in the intervention and comparison communities then this may not matter a great deal. However, if in-migration rates are high and/or differ greatly between the intervention and comparison communities then this could matter for our impact estimates.

A. Study design

We plan to conduct three types of analyses to try to address issues related to migration. First, we will estimate impacts of the T&D and FS components of the energy sector project on migration rates into and out of the study communities. Second, we will look at changes in outcomes for in-migrants ex-post. Finally, we will estimate impacts on property values, which in theory could help to capture impacts that might occur via migration.

Migration rates. We plan to estimate impacts on the total population, in-migration rates, out-migration rates, and new household formation rates.[39] We can do this using three data sets—the community survey, the household listing, and the household survey data. The community survey data will enable us to capture impacts for larger communities (villages as opposed to the subvillages in rural areas) while the household listing data should give us more accurate measures that are focused on the subvillages where the largest impacts were expected (those subvillages selected for the household surveys). These first two data sets will enable us to estimate impacts on all of the migration-related outcomes described above. The household survey data will only enable us to estimate impacts on out-migration. However, the household survey data will enable us to see how out-migration varies with baseline household characteristics.[40] When doing these analyses,

[39] New household formation refers to households that are found in the follow-up survey that have heads who were non-heads in other households in the same community at baseline. This could happen if, for example, a young adult moves out of his or her parents' home and forms a new household. These households are included in our household listing and we will survey a sample of them as part of our "in-migrant" sample.

[40] The household listing data will also enable us to do this, but with a very limited set of household characteristics—if within 30 meters of an existing line and if connected.

we will control for baseline community-level characteristics as well as household-level characteristics.

If the new lines are affecting migration rates then we might expect to see households migrating into areas closer to the new lines within a community; if new lines reduce out-migration, we might expect impacts on out-migration to also be more pronounced in the areas closer to the new lines. To test these theories, we will calculate the percentage change in the population within 30 meters of the new lines compared to the percentage change in the population in the rest of the communities receiving the new lines.[41] We will only be able to do this in the communities that receive new lines; therefore, these will not be impact estimates; nevertheless, they will provide useful information about the possible role of the new lines in changing migration rates.

Changes in outcomes for migrants. The household listing will cover all households in these communities at the time of the follow-up survey, including both in-migrants and newly formed households. We will sample a random subset of the in-migrants, question them about their pre-migration status, and calculate how their outcomes changed after migration. We can also compare these changes to changes observed for households that did not migrate that were similar at baseline. These will be non-rigorous descriptive comparisons because we will be relying on retrospective data for the in-migrants. We will also do a closely related analysis for internal migrants—those households that migrated within their communities (including newly formed households, some of which will also be included in our household survey). For those internal migrants, we will be able to estimate how outcomes changed for those who moved to be within 30 meters of the new lines, compared to outcomes for similar households that did not migrate at all or that did not migrate to within 30 meters of the new lines. These will also be descriptive analyses and will exclude in-migrants from other communities, but will have the benefit of using baseline data for both the migrant/newly formed household sample and for the comparison sample of non-migrants.

We plan to sample 2,000 in-migrants and newly formed households. This will enable us to estimate changes in the use of solid fuel as small as around 13 kg. This is based on assumptions similar to those used to estimate the MDIs for the impacts reported in Chapter II.[42]

Property values. Looking at property values may enable us to capture benefits not captured by many other household-level outcomes. The benefit of looking at household-level outcomes is that it enables us to estimate benefits for both owners and non-owners. However, it will not allow us to capture benefits that go to landlords who live in other communities. This matters because some landlords may raise rent or sell their land to in-migrants when electricity arrives, meaning

[41] This analysis may face some challenges because the GPS data at baseline appears to have had some problems. Consequently we may need to conduct the analysis for a subset of the baseline communities—those that appear to have good baseline GPS data as confirmed by checking those data against the follow-up GPS data to see if they appear to be in the correct community. We know that at least some of the baseline GPS data were bad because they show the communities as being outside of the study regions. We are not sure how much of the data that appear within the study regions are good.

[42] There are two differences. First, there is no comparison group, which reduces the MDI. Second, the sample size is much smaller, which increases the MDI. The net effect is a slight reduction. This is also based on an assumed 90 percent response rate.

that the current tenants may not benefit as much as the landlords.[43] To help capture these potential benefits of electricity, we propose to look at property values as an outcome in our analyses. The energy sector project may impact property values if at least some households in a community consider selling their land and the prices they are offered change because of the new lines. This may not happen if land sales are very limited and/or households do not have a good sense of how much more their land should be worth because of the new lines. Consequently, we will have to assess the quality of the data on property values reported by the respondents before estimating impacts on them.

Our ability to estimate impacts of the T&D activity and FS initiative on property values is limited by the fact that households may choose to improve their properties when electricity arrives. Consequently, it may be difficult to disentangle the impacts of the new lines from the impacts of other changes that are made concurrently. We will try to disentangle these impacts by controlling for other characteristics of the property, such as size of the house, materials, and related amenities (namely plumbing and toilet types).

In the same way that we might expect the impacts of the project on migration to be larger within 30 meters of the new lines, we might also expect impacts on property values to be larger there. For this reason, we will test to see if property values rose more for households within 30 meters of the new lines than for other households in the same communities that were not as close to the new lines. When estimating impacts on property values, we will use methods similar to those used for the other household outcomes described in Chapter II. The major difference will be that we will use property values reported by in-migrants in the households formerly occupied by out-migrants.

B. Data collection

We will collect most of the data on migration-related outcomes through the surveys we are conducting for the rest of our analyses (the community survey, household listing, and household survey following the baseline survey sample). In addition, however, we added a supplement to our household survey to include up to 2,000 households that moved into the study communities since 2011. We will ask these in-migrants questions about their life in 2011 so that we can capture changes over time for these households and compare those changes to changes for other households.

[43] Indeed, some baseline tenants may be forced to move by the increased rents and, thus, be negatively impacted by the new lines. We should be able to capture negative impacts on these households by following the out-migrants.

This page has been left blank for double-sided copying.

V. EVALUATION OF ZANZIBAR INTERCONNECTOR ACTIVITY

Our report on the Zanzibar interconnector activity is now complete (Schurrer et al. 2015). This chapter describes the research questions addressed in that study, how we collected the data, and our analytic approach.

A. Research questions

We designed the evaluation of the cable activity in close cooperation with MCC, MCA-T, and stakeholders in Zanzibar. The hotel study addressed the following two research questions:

1. How did outcomes for large hotels on Unguja Island change after the new submarine cable was installed compared to outcomes before cable installation?
2. What are the large hotel owners and managers' perceptions of changes in their hotel outcomes after the cable was installed?

We answered the first question using a pre-post analysis of data from repeated surveys of 30 large hotels on Unguja Island. We addressed the second question using qualitative data collected from the hotel owners and managers who responded to the post-cable hotel survey.

B. Data sources

The 30 hotels selected for the study were sampled from a list of hotels on Unguja Island. The list contained 306 hotels (based on 2008 data) and was provided by the Zanzibar Commission for Tourism. For purposes of the study, hotels were required to have (1) 10 or more rooms and (2) an international grade of at least one star. This helped us to identify the larger and higher quality hotels that might be more likely to make use of electricity. After applying those restrictions, 45 hotels remained; we selected a random subset of those hotels.[44]

Mathematica's data collection partner for the Zanzibar baseline (DHI International) surveyed the selected 30 hotels three times each before the cable activity began, in the months of June, July, and August 2010. The survey contained mostly quantitative questions about the reliability and quality of electricity the hotels were receiving, and their revenues and costs that might be affected by electricity. CSR Group Africa, Mathematica's data collection partner for the follow-up hotel study (as well as for the qualitative study), resurveyed these hotels—again three times each, in July, August, and September 2014—about 15 months after the new cable was inaugurated.[45] Two hotels did not respond to the post-cable survey; therefore, we ended up with a total sample size of 28 hotels with data for both the pre- and post-cable periods.[46] Typically, managers or finance

[44] Hotels that were out of business, not connected to ZECO's electricity network, or that refused to participate in the survey were replaced by other randomly selected hotels.

[45] We planned to obtain data for the same months in each period, but the data collection was delayed at follow-up due to problems obtaining clearance from the local government in Zanzibar.

[46] The 2 hotels that did not respond were similar to the other 28 hotels in terms of the average number of staff and number of rooms; however, they had lower revenues but higher energy costs.

department staff responded to the surveys, which were administered in person by the data collection staff.

The surveys contained questions about a variety of hotel outcomes and were designed to capture variation across months on relevant questions. The questionnaire contained quantitative and qualitative questions about the reliability and quality of electricity hotels were receiving, as well as their revenues and costs that might have been affected by electricity during the 30 days preceding the survey.[47] The survey instrument was pilot tested in May 2010 and revised to incorporate lessons learned during the pilot. The first month of both the pre-cable baseline and post-cable follow-up surveys asked questions about hotel characteristics that were not asked in subsequent waves, such as number of rooms and staff.

C. Outcomes and analytic approach

For our analysis, we identified five domains of hotel outcomes: energy costs, revenue, quality of electricity, hotel staffing, and guest satisfaction (Table V.1). These domains and their associated outcomes map to the short- and intermediate-term results described in the conceptual framework in Chapter I of Schurrer et al. (2015). These are the outcomes for which we have data from the pre- and post-cable surveys and that we can reasonably expect to have been affected by the cable activity at the time of the post-cable survey.

Table V.1. Hotel outcomes by domain

Domain	Outcomes
Revenue	• Monthly revenue from lodging (adjusted for inflation)[a] • Occupancy rate • Room rate • Number of rooms
Energy costs	• Total energy costs (adjusted for inflation) • Grid electricity cost • Generator, diesel, and other costs[b] • Electrical device repair and replacement costs • Presence, type, and number of electrical devices like TVs and air conditioners
Quality of electricity	• Number of outages • Number of voltage fluctuations
Staffing	• Total number of staff
Guest satisfaction	• Satisfaction rate[c]

[a] Monthly revenue from lodging is the product of the occupancy rate, room rate, and number of days in the month.

[b] Other costs include kerosene, bottled gas, charcoal, firewood, solar power, batteries, candles, and flashlights.

[c] The satisfaction rate is the percentage of hotels not reporting a cancellation, shortened stay, or complaint due to electricity problems.

[47] The June 2010 survey also contained qualitative questions aimed at helping respondents describe the experiences of the hotels during the blackout that lasted from December 2009 to March 2010 on Unguja and the various coping strategies utilized. The Zanzibar baseline report (Hankinson et al. 2011) contains findings from these data.

When estimating the differences in hotel outcomes before and after the new cable, we used regression models with hotel fixed effects to account for time-invariant characteristics of the hotels. Because the pre-cable and post-cable surveys were administered during a different set of months, we have two analytic month sets: overlapping months (July and August in both years) and non-overlapping ones (June 2010 and September 2014). We specify two regression models to account for the different months in which the survey was fielded in the pre- and post-cable periods. The first regression model uses an equation of the following form that estimates the differences in hotel outcomes using hotel survey data for all months:

(1) $Y_{it} = \beta_0 + \beta_1 POST_t + \alpha_t + \delta_i + e_{it}$

where

Y_{it} is the outcome for hotel i in month t,

β_0 is the intercept,

$POST_t$ is a dummy variable identifying periods after the cable is completed,

β_1 is an estimate of the pre-post difference in outcome Y_{it},

α_t are fixed-month effects for July and August, assumed constant across years,

δ_i are hotel fixed effects, also assumed to be constant across years, and

e_{it} is the error term.

To examine the influence of the non-overlapping months on the estimates, we use the following equation:

(2) $Y_{it} = \beta_0 + \beta_2 POST_t + \pi POST_t * OTHER_t + \alpha_t + \delta_i + e_{it}$

where

β_2 is an estimate of the effect of the cable activity in the overlapping months (July and August),

$OTHER_t$ is a dummy for the non-overlapping months (June 2010 and September 2014), and

π is the estimate of the effect of the cable activity in the non-overlapping months (June 2010 and September 2014).

In the main body of the report, we focused on results from the overlapping months. The differences in estimated outcomes using all months may be biased due to differences in June and September. However, the results are consistent across both models for most outcomes and we noted when the estimated effects are sensitive to model specification.

A bias could also arise in the estimated differences based on July and August data, since some hotels did not respond to all items in the post-cable survey. To help ensure that the July/August data provided balanced comparisons of measures across years, we used data for hotels that responded to the relevant item at both baseline and follow-up rounds of the survey. This was done separately for each outcome so that all analyses based on overlapping months maintain this balance.

We used a price deflator to convert all outcomes measured in U.S. dollars into constant December 2014 dollars. Specifically, we used the National Consumer Price Index, available from the National Bureau of Statistics, Tanzania, to adjust the dollar amounts. Before applying the inflation adjustments, we converted all values reported in Tanzanian shillings (TZS) to U.S. dollars using an exchange rate of TZS 1,411 = US $1 in 2010, and TZS 1,662 = US $1 in 2014.

The first wave of the post-cable survey also contained qualitative questions designed to probe hotel staff on the effects of the cable activity on revenue, staffing, and operations. We reviewed the discussion transcripts and identified major themes that emerged across respondents. These themes centered on energy quality, increased competition in the hotel sector, the price of electricity, and views on ZECO’s customer service.

As noted above, results from these analyses are presented in Schurrer et al. (2015).

VI. TIMELINE AND TASKS AHEAD

In this chapter, we describe the overall energy sector evaluation timeline and the tasks ahead. These cover the evaluation of the T&D activity, the FS initiative, and the Kigoma solar activity. The evaluation of the Zanzibar cable activity has been completed (as discussed in Chapter V), but for completeness, we include the timeline for that evaluation here as well.

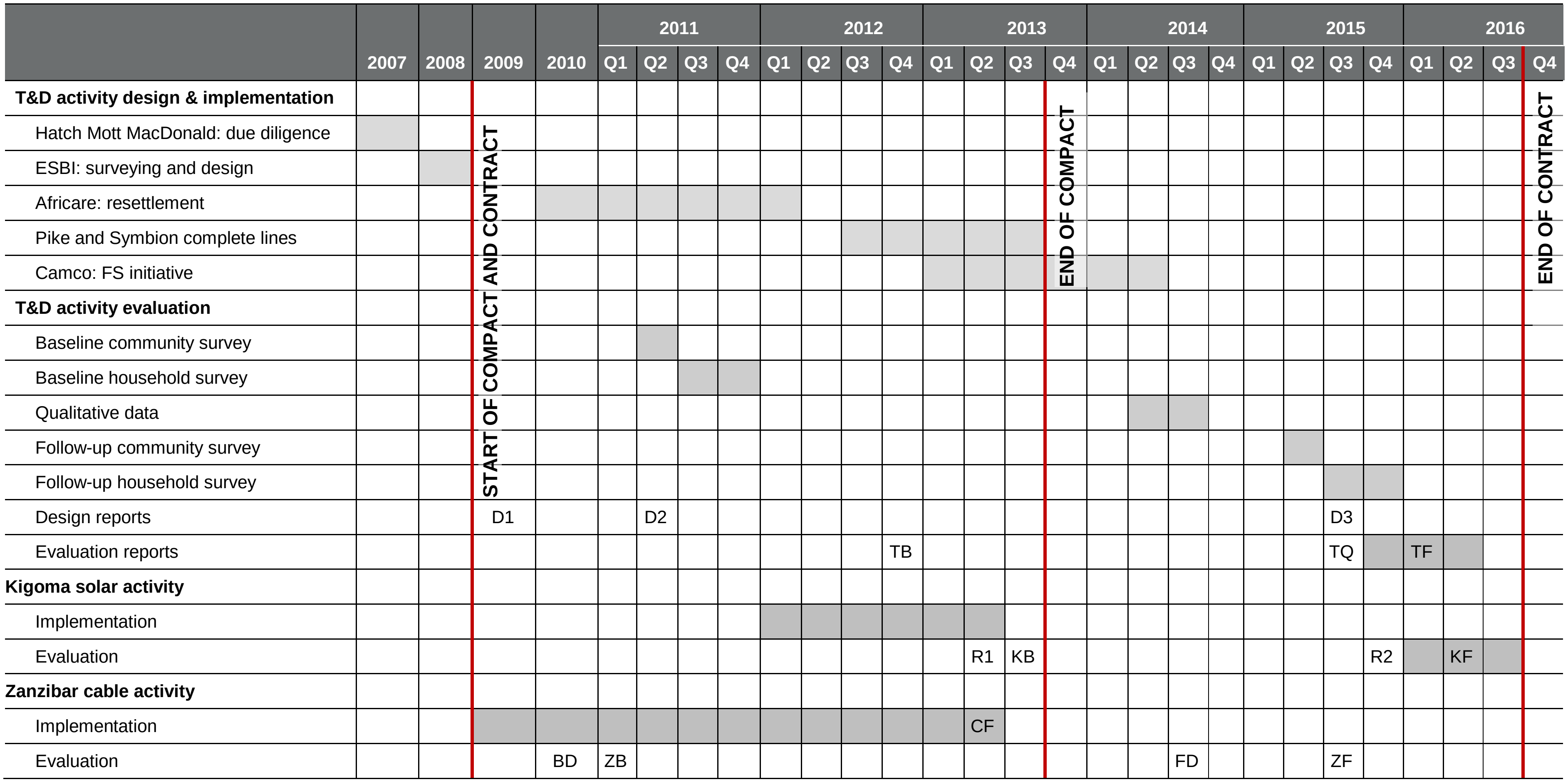
A. Timeline

Figure VI.1 presents the actual and planned timelines for the energy sector project implementation and evaluation. In 2007, MCC hired the engineering firm Hatch Mott and McDonald to produce a preliminary design for the T&D component of the energy sector project. Then, in 2008, MCC hired the engineering firm ESBI to finalize the design and to oversee the contractors (Pike and Symbion) that built the new T&D lines. Mathematica was hired to start work on the evaluation in July 2008 and the compact itself began in September of that year. Much of 2008 and 2009 was devoted to implementation and evaluation planning. In 2010 and 2011, Africare helped with resettlement of households that needed to move in order to make way for the new lines. Lines were installed in some communities by third quarter 2012 but others were not completed until close to the compact's end, in September 2013. The FS initiative followed on after completion of lines within each community and, thus, was delayed by a few months, with connections in some communities completed as late as mid-2014.

For the evaluation of the T&D activity and the FS initiative, we completed an early design memorandum in 2009 and updated it with a design report in 2011. We completed the baseline household and community surveys in 2011, well before any communities had new lines. We presented findings from the analysis of baseline survey data in a baseline report in 2012. In 2014, shortly after completion of the FS initiative, we collected qualitative data on implementation and perceived benefits of electrification, to understand the activity's effects in the interim between surveys. Our report on these data is undergoing final approval and should be released shortly. We are in the process of collecting the follow-up community survey data and will conduct follow-up surveys beginning in September 2015, well after the completion of both the T&D lines and FS work. This final design report builds on both the earlier design report and the baseline report. Because of its timing, it is not influenced by the follow-up household survey, and only minimally by data from the follow-up community survey. In addition, as discussed earlier, we are making only minor changes to our plans for the T&D and FS impact analysis. The major changes involve adding migration as a key outcome and clarifying the other key outcomes.

Implementation of the Kigoma solar activity was expected to occur between March 2012 and May 2013, but the first round of data were not collected until July 2013, after about two-thirds of the activity had been implemented. Thus, we lack true baseline information for most components of this activity. We do have retrospective data for those components, however. The round one data have true baseline information for one component—the distribution of hospital refrigerators—as it took place after the first round of data collection. We plan to collect a second round of survey data for the Kigoma solar evaluation in October 2015. This will contain true follow-up information for all beneficiaries, because the solar intervention ended in 2014.

Figure VI.1. Timeline for the energy sector project implementation and evaluation



Notes: The rows labeled "Pike and Symbion complete lines" and "Camco: FS initiative" show the range of completion dates by community for the new T&D lines and FS initiative, respectively; implementation of each component started earlier.
Q1 = January–March, Q2 = April–June, Q3 = July–September, and Q4 = October–December.
D1 = Design report 1 finished, D2 = Design report 2 released, D3 = First draft design report 3.
TB = T&D Baseline report released, TQ = T&D Interim qualitative report released, TF = First draft of final T&D impact report.
BD=Baseline data, FD = Follow-up data, R1 = Round 1 data, R2 = Round 2 data,
ZB = Zanzibar baseline report released, CF = New cable finished, ZF = Zanzibar final report released.
KB = Kigoma baseline report released, KF = First draft Kigoma final report.

For the Zanzibar cable activity, the installation of the submarine interconnector cable and the rehabilitation of substation and distribution lines on the Unguja Island were completed by June 2013. For the evaluation of the activity, we surveyed the selected hotels three times each, in June, July, and August 2010, before the cable activity was implemented. We conducted a follow-up survey with these hotels—again three times each, in July, August, and September 2014. We present the findings from the analysis of the hotel survey data in the Zanzibar final report (Schurrer et al. 2015), which is in the process of being approved for release.

B. Tasks ahead

As noted above, the final Zanzibar report and the qualitative report are currently undergoing review prior to release. In addition, the community survey data collection for the T&D evaluation is underway. We will use those data to determine which households to survey in the fall and whether or not to further postpone the household survey. We plan to survey up to 2,000 in-migrant households. We will survey all in-migrants if there are fewer than 2,000 and a random subset if there are more than 2,000. In addition, if there are fewer than 1,000 connected households near to the new lines covered by our baseline household survey data, we will either add additional households from the baseline listing that did end up connecting to the new lines to ensure that we cover at least 1,000 connected households in our final analysis of non-migrant households or else postpone the household survey to allow more time for households to connect.[48] The decision to postpone the household survey would be based on an estimate of how much the difference in connection rates was likely to change in future months. Thus, a higher increase in connection rates in the intervention areas would suggest that we might recommend further postponing the household survey. On the other hand, if connection rates are growing by even more in the comparison group then we probably would not want to postpone the household survey.

We are currently finalizing the household survey and Kigoma solar survey instruments. Data collection for those surveys will commence in late August or early September 2015 and will finish by December 2015.

We will analyze the data from the follow-up surveys and prepare reports to present the findings. We plan to complete the first draft of the final T&D and FS impact report by March 2016, and the first draft of the Kigoma solar final report by April 2016. It may take a few months after that for these final reports to be released, given the need for multiple rounds of stakeholder reviews and approval by MCC and the government of Tanzania.

[48] If this happens, it would suggest that the intervention had fairly small impacts on connection rates by the time of the community survey. In that case, we would likely find fairly small overall impacts but might still be able to detect impacts of getting electrified using the oversample of electrified households. If we oversample electrified households, then we would undersample other households to maintain the same overall sample size, and reweight the sample to account for this oversampling when estimating impacts of the T&D and FS initiative (Manski 1999).

This page has been left blank for double-sided copying

REFERENCES

Angrist, Joshua, and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Chapter 3: "Making Regression Make Sense." Princeton, NJ: Princeton University Press, 2009.

Angrist, Joshua, Guido Imbens, and Don Rubin. "Identification of Causal Effects Using Instrumental Variables." JASA Applications invited paper, with comments and authors' response. *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 444–55.

Barnes, Douglas, and Liu Qian. "Urban Interfuel Substitution, Energy Use, and Equity in Developing Countries: Some Preliminary Results." Working Paper, Energy Series, no. 53. Washington, DC: The World Bank, Industry and Energy Department, 1992.

Barnes, Douglas. "Electric Power for Rural Growth: How Electricity Affects Rural Life in Developing Countries." *Rural Sociology*, vol. 54, no. 1, 1988.

Bernard, Tanguy, and Maximo Torero. "Welfare Impact of Rural Electrification: 'Short Term' Evidence from Ethiopia: A Report for the Ethiopian Electric Power Corporation and the World Bank." Washington, DC: International Food Policy Research Institute, December 2009.

Bloom, Howard S. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review*, vol. 8, no. 2, 1984, pp. 225–46.

Busalama, Abel Y. "Kigoma Solar Baseline and Interim Performance Evaluation Report." Submitted to the Millennium Challenge Account–Tanzania, 2013.

Chaplin, Duncan, Arif Mamun, and John Schurrer. "Evaluation of the Millennium Challenge Corporation's Electricity-Transmission and Distribution Line-Extension Activity in Tanzania: Baseline Report." Submitted to the Millennium Challenge Corporation. Washington, DC: Mathematica Policy Research, November, 2012.

Chaplin, Duncan, Arif Mamun, Thomas Fraker, Kathy Buek, Minki Chatterji, and Denzel Hankinson, "Evaluation of Tanzania Energy Sector Project: Updated Design Report." Washington, DC: Mathematica Policy Research, March 16, 2011.

Chaplin, Duncan, and Arif Mamun. "Quality of Data on Connections to the National Electric Grid for Evaluation of Millennium Challenge Corporation's Energy Project in Mainland Tanzania." Draft memorandum submitted to the Millennium Challenge Corporation. Washington, DC: Mathematica Policy Research, June 10, 2015.

Collings, Simon. "Phone Charging Micro-businesses in Tanzania and Uganda," Global Village Energy Partnership International, 2011. Available at http://www.gvepinternational.org/sites/default/files/phone_charging_businesses_report_with_gsma_final_for_web_0.pdf. Accessed September 25, 2012.

Davis, Mark, “Rural Household Energy Consumption: The Effects of Access to Electricity—Evidence from South Africa.” *Energy Policy*, vol. 26, no. 3, February 1998, pp. 207–217.

Dinkelman, Taryn. “The Effects of Rural Electrification on Employment: New Evidence from South Africa.” *American Economic Review*, vol. 101, no. 7, December 2011, pp. 3078–3108.

Eberhard, Anton, Vivien Foster, Cecilia Briceño-Garmendia, Fatimata Ouedraogo, Daniel Camos, and Maria Shkaratan. “Underpowered: The State of the Power Sector in Sub-Saharan Africa.” Background Paper 6, Africa Infrastructure Country Diagnostic. Washington, DC: The World Bank, 2008.

Energy and Environment Partnership. “Tanzania” (nd). Available at http://www.eepafrica.org/index.php?option=com_content&view=article&id=83:tanzania&catid=904:profiles&Itemid=80. Accessed September 25, 2012.

Energy Sector Management Assistance Program (ESMAP). “Technical and Economic Assessment of Off-Grid, Mini-Grid and Grid Electrification Technologies.” ESMAP Technical Paper 121/07. Washington, DC: The World Bank, December 2007a.

Energy Sector Management Assistance Program (ESMAP). “Maximisation des Retombées de l’Electricité en Zones Rurales, Application au Cas du Sénégal.” ESMAP Technical Paper 109/07 FR. Washington, D.C.: World Bank, 2007b.

Engelke, Courtenay. “Improving Air Quality Through Energy Efficiency in Mongolia.” Knowledge & Innovation Network, *The Journal of MCC&MCA Technical Staff*, vol. 1, no. 1, summer/fall 2012.

Heckman, James, Hidehiko Ichimura, and Petra Todd. “Matching as an Econometric Estimator.” *Review of Economic Studies*, vol. 65, 1998, pp. 264–294.

Heltberg, Rasmus. “Household Fuel and Energy Use in Developing Countries: A Multicountry Study.” Oil and Gas Policy Division, World Bank. Washington, DC: The World Bank, 2003.

Hosier, R., and W. Kipondya. “Urban Household Energy Use in Tanzania: Prices, Substitutes, and Poverty.” *Energy Policy*, vol. 21, no. 5, 1993, pp. 454–473.

Imai, Kosuke, and David van Dyk. “Causal Inference with General Treatment Regimes: Generalizing the Propensity Score.” *Journal of the American Statistical Association*, vol. 99, no. 467, September 2004.

International Renewable Energy Agency (IRENA). “Renewable Energy Technologies: Cost Analysis Series: Solar Photovoltaics.” IRENA Working Paper Series, vol. 1, no. 4/5, June 2012.

Jacobson, Arne. “Connective Power: Solar Electrification, and Social Change in Kenya.” *World Development*, vol. 35, no. 1, 2007, pp. 144–162.

Ketlogetswe, C., T.H. Mothudi, and J. Mothibi. “Effectiveness of Botswana’s Policy on Rural Electrification.” *Energy Policy*, vol. 35, no. 2, 2007.

Khandker, Shahidur, Douglas Barnes, Hussain Samad, and Nguyen Huu Minh. “Welfare Impacts of Rural Electrification: Evidence from Vietnam.” Policy Research Working Paper no. 5057. Washington, DC: The World Bank, September 2009.

Lohr, Sharon L. *Sampling: Design and Analysis*. Duxbury Press, 1999.

Madubansi, M., and C.M. Shackleton, “Changing Energy Profiles and Consumption Patterns Following Electrification in Five Rural Villages, South Africa.” *Energy Policy*, vol. 34, 2006, pp. 4081–92.

Manski, Charles F. “Nonparametric Identification Under Response-Based Sampling,” 1999. In *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, edited by C. Hsiao, K. Morimune, and J. Powell. New York: Cambridge University Press (in press).

Martin, Derrel, Tom Dorn, Steve Melvin, Alan Corr, and William Kranz. “Evaluating Energy Use for Pumping Irrigation Water.” Proceedings of the 23rd Annual Central Plains Irrigation Conference, Burlington, CO, February 22–23, 2011.

Martins, Johan. “The Impact of the Use of Energy Sources on the Quality of Life of Poor Communities.” *Social Indicators Research*, vol. 72, 2005, pp. 373–402.

Millennium Challenge Account–Tanzania (MCA-T). “Tanzania Energy Sector Project Logic.” Internal monitoring and evaluation documentation. Dar es Salaam, Tanzania: Millennium Challenge Account–Tanzania, January 19, 2012.

Millennium Challenge Corporation (MCC). “Millennium Challenge Compact between the United States of America, acting through the Millennium Challenge Corporation, and the Government of the United Republic of Tanzania, acting through the Ministry of Finance.” Washington, DC, February 2008.

Miller, Candace, John Schurrer, Nicholas Redel, Arif Mamun, and Duncan Chaplin. “Millennium Challenge Corporation’s Energy Project in Mainland Tanzania: Qualitative Evaluation.” Draft report submitted to the Millennium Challenge Corporation. Washington, DC: Mathematica Policy Research, March 4, 2015.

Ministry of Finance and Economic Affairs, Government of Tanzania (MoF). “United Republic of Tanzania: Poverty and Human Development Report 2009.” Dar es Salaam, Tanzania: MoF, 2009.

Ministry of Finance and Economic Affairs, Revolutionary Government of Zanzibar (Zanzibar MoF). “Zanzibar Poverty and Human Development Report 2009.” Zanzibar, Tanzania: Zanzibar MoF, 2009.

Moss, Todd, “The Next Administration Should Close Africa’s Energy Poverty Gap.” Center for Global Development blog, Rethinking U.S. Foreign Assistance. Washington, DC, September 24, 2012.

National Bureau of Statistics, Tanzania. “Household Budget Survey 2007—Tanzania Mainland,” 2009.

National Bureau of Statistics (NBS) and ICF Macro. "2010 Tanzania Demographic and Health Survey." Calverton, MD: Tanzania NBS and ICF Macro, 2011.

Paster, Pablo. "Ask Pablo: Electric Kettle, Stove, or Microwave Oven?" Technology/Clean Technology Section at TreeHugger, Discovery Communications, LLC, November 23, 2009. Available at http://www.treehugger.com/clean-technology/ask-pablo-electric-kettle-stove-or-microwave-oven.html. Accessed October 15, 2012.

Pfefferman, D. "The Role of Sampling Weights When Modeling Survey Data." *International Statistical Review*, vol. 61, 1993, pp. 317–37.

Robins, J., and A. Rotnitzky. Comment on "Inference for Semiparametric Models: Some Questions and an Answer," by J. P. Bickel and J. Kwon. *Statistica Sinica*, vol. 11, 2001, pp. 920–936.

Rosenbaum, Paul R., and Donald B. Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, vol. 70, 1983, pp. 41–55.

Rubin, Donald. "The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies." *Biometrics*, vol. 29, no. 1, March 1973, pp. 185–203.

Rubin, Donald. "The Design Versus the Analysis of Observational Studies for Causal Effects: Parallels with the Design of Randomized Trials." *Statistics in Medicine*, vol. 26, 2007, pp. 20–36.

Schochet, Peter Z. "An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations." *Education Review*, vol. 33, no. 6, December 2009.

Schurrer, John, Nicholas Redel, Duncan Chaplin, and Arif Mamun. "Evaluation of the Zanzibar Interconnector Activity: Findings from the Hotel Study." Draft report submitted to Millennium Challenge Corporation. Washington, DC: Mathematica Policy Research, May 6, 2015.

Silverman, B. W. Density Estimation for Statistics and Data Analysis. London: Chapman and Hall, 1986.

The Economist. "Not Such a Bright Idea: Making Lighting More Efficient Could Increase Energy Use, Not Decrease It," August 26, 2010. Available at http://www.economist.com/node/16886228. Accessed Sept. 27, 2012.

United States Agency for International Development (USAID). "Power Africa: Annual Report 2014." Washington, DC: USAID, July 2014. Available at http://www.usaid.gov/sites/default/files/documents/1860/USAID_PowerAfrica_AR_July2014.pdf.

United Nations. "Explanatory Note on the 2014 Human Development Report Composite Indices, Tanzania (United Republic of)," United Nations Development Programme, 2014. Available at http://hdr.undp.org/sites/all/themes/hdr_theme/country-notes/TZA.pdf. Accessed June 25, 2015.

Woofenden, Ian, "Off or On Grid?," *Home Power*, January 2012. Available at https://homepower.com/articles/or-grid. Accessed September 24, 2012.

World Bank. "World Bank Indicators." Available at http://data.worldbank.org/country/tanzania. Accessed August 15, 2014.

World Bank. "World Development Indicators 2010." Washington, DC: The World Bank, 2010.

World Bank. "The Welfare Impact of Rural Electrification: A Reassessment of the Costs and Benefits," an independent evaluation group impact evaluation. Washington, DC: The World Bank, 2008.

This page has been left blank for double-sided copying.

APPENDIX

SAMPLING AND MATCHING WEIGHTS FOR THE T&D EVALUATION

This page has been left blank for double-sided copying.

In this appendix we describe how we created weights for the T&D evaluation.

1. Nonresponse sampling weights for intervention group

For our intervention group, we created weights to adjust for sampling and survey nonresponse. Households in the intervention group were sampled based on approximate eligibility for a subsidy pilot intervention that was later replaced by the financing scheme. Approximate eligibility was based on whether or not the household appeared to have two or fewer rooms. The survey team made this determination during the household listing process in the intervention areas. They then oversampled those households so that 40 percent of the resulting sample qualified, compared to 25 percent in the sampling frame. We created sampling weights to adjust our sample to be representative of the full population in the intervention group. These sampling weights (SW_i) were calculated as one over the probability of being sampled.

$$SW_i = \frac{1}{Pr_i} \text{ where } Pr_i = \text{probability household i was sampled.}$$

We then adjusted these sampling weights for nonresponse using 18 categories for nonresponse. These categories were based on region and total migration (in-migration plus out-migration as reported in the community survey). First we created three categories for total migration. Then we calculated the response rate for each of these categories by region (R_i). Lastly, we multiplied the sampling weights by the inverse of response rates to create a final weight for the intervention group (W_i).

$$W_i = \frac{SW_i}{R_i}$$

We also created weights for the comparison group to be used for pre-match comparisons. Consequently, the comparison group weights adjust for nonresponse by community but not for sampling since all households were sampled with equal probability within a community.

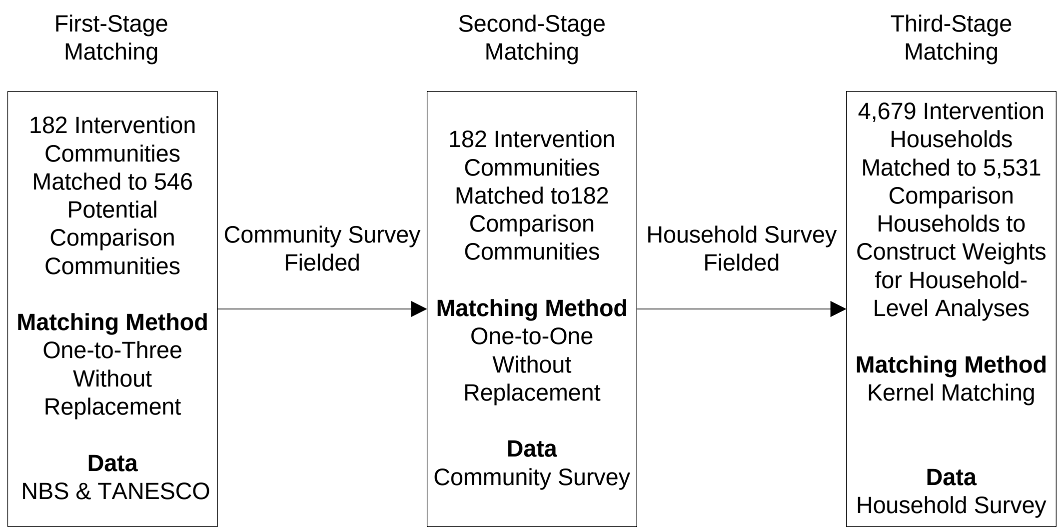
Choosing the number of households to sample in each comparison community was nontrivial. For the intervention group it was easy because we had the household listing long before we did the household survey. For the comparison group, however, we did not have the listing until the day before the household survey was done. Moreover, when we collected the household listing for the intervention group we learned that the community survey reports on community size were not always accurate. Consequently, we adjusted the community survey responses for the comparison group to obtain a better estimate of the number of eligible households. More precisely, we used the household listing data in the intervention group and regressed the number of eligible households in the community on the number reported in the community survey and other community characteristics. We then used the coefficients from this regression to create predicted community size variables for the comparison group.

2. Propensity score matching weights for comparison group

Our initial set of 182 comparison communities was selected through two stages of propensity score matching. The first stage used data from the National Bureau of Statistics (NBS) and the

Tanzania Electricity Supply Company (TANESCO) to identify a set of 546 potential comparison communities. The community survey was fielded in these communities and the data were used to identify the 182 matched communities in the second stage of matching using a nearest neighbor algorithm. The household survey was administered in 182 comparison and 182 intervention communities.[49] For this baseline report, we conducted a third stage of matching using household survey data to produce household weights that further improve the quality of our matches across important characteristics and enhance our future ability to make inferences about the impact of the T&D extension on household and community outcomes. We use a kernel matching method to construct a set of matched sample weights WM for the comparison group so that the weighted average of their outcomes could serve as a defensible counterfactual for those of the intervention group. Figure A.1 presents the three stages of matching. The remainder of this section describes the methodology used in this third stage of the propensity score matching and weight construction process.

Figure A.1. Stages of matching used to identify comparison communities and households



a. Estimation of the propensity score

The first step in the construction of the matched sample weights was the estimation of a logistic regression model, where the dependent variable $Interv_i$, indicating whether household i was a member of the intervention sample, was regressed on a 1xk vector of baseline characteristics X_i:

[49] Seven of the original 182 comparison communities were replaced during fielding of the household survey because of a lack of eligible households (NRECA 2012, Table 5). The community in the set of 364 unmatched communities with the closest propensity score to the original matched community was selected as the replacement.

$$\Pr(Interv_i = 1) = \Lambda(\mathbf{X}_i\boldsymbol{\gamma}) = \frac{\exp(\mathbf{X}_i\boldsymbol{\gamma})}{1+\exp(\mathbf{X}_i\boldsymbol{\gamma})}, \quad (1)$$

where γ is a *kx1* parameter vector.

To estimate (1), we weighted each intervention household by the nonresponse adjusted sample weight (described earlier in this appendix) from the household survey, W_i, and set the weights for the comparison group to one. From the estimation results, we obtained each comparison and intervention household's estimated propensity score as the predicted probability, $\hat{q} = \Lambda(\mathbf{X}_i\hat{\boldsymbol{\gamma}})$, of belonging to the intervention sample.

A critical methodological challenge for propensity score analysis is specifying a model that satisfies two important criteria. First, the model should include important observable characteristics that are likely correlated with the outcomes of interest, and predict membership in the intervention group. Second, the model needs to satisfy the balancing property in order to make inferences about the effect of the intervention on the outcomes (Rosenbaum and Rubin 1983). In theory, this means that for every value of the propensity score, there is no statistically significant difference between the intervention and comparison groups for the matching variables used to estimate the propensity score. In practice, the observations are divided into several blocks based on their propensity scores, and the balancing property is satisfied when there are no statistically significant differences between the intervention and comparison groups for the matching variables within each block.

To satisfy these two criteria, we iterated though a series of models that included household-level variables (1) thought to be correlated with characteristics that predict access to electricity and (2) with significant differences between intervention and comparison households. This covered many of our key outcomes related to income and energy. We also included gender of the household head, given the interest in gender differences. We started with a limited set of variables, performed the matching, and tested for post-match differences across a larger set of variables. We then respecified the propensity score model, including variables that still had post-match differences that were statistically significant. Our final model satisfied the balancing property described above in all seven propensity score blocks and produced a sample that was balanced overall for the larger set of variables, as discussed in Chapter IV. The final propensity score regression included the following variables:

- Gender of the household head
- Household moved in the last 7.5 months
- TV hours per month
- Presence of any phone (mobile or landline)
- Total number of appliances
- House has an electrifiable roof
- Number of rooms in the house. Constructed three binary variables based on the distribution of the number of rooms:

   - Zero to two rooms (minimum to 25th percentile)
   - Two rooms (25th to 50th percentile)
   - Three to 20 rooms (50th percentile to maximum)

- Annual consumption (TZS)
- Total annual income (TZS)
- Total assets (TZS)
- Electricity expenditures per year (TZS)
- Electricity expenditures per year squared (TZS)
- Total amount spent on energy per year (TZS). Constructed four binary variables based on the distribution of the amount spent on energy:
   - 0–90,000 (minimum to 50th percentile),
   - 90,001–480,000 (50th to 75th percentile),
   - 480,001–840,000 (75th to 99th percentile), and
   - 840,000–3,204,000 (99th percentile to maximum)
- kWh per month from the electrical grid
- Non-electric energy made per month, including from small batteries (kWh)

After estimating the propensity score, we determined that there was sufficient overlap of the propensity scores between the intervention and comparison households to proceed with the kernel matching and creation of matched sample weights (described below). Table A.1 shows the summary statistics of the propensity scores by intervention status for the full set of intervention and comparison group households. Although the means of the propensity score were similar, the difference between the intervention and comparison group propensity score means was statistically significant (t = -14.33), indicating that the two groups differ before applying the matching weights (as expected). Table A.1 suggests that the mean difference was in part due to the difference in the 25th percentile and in the 90th percentile and above.

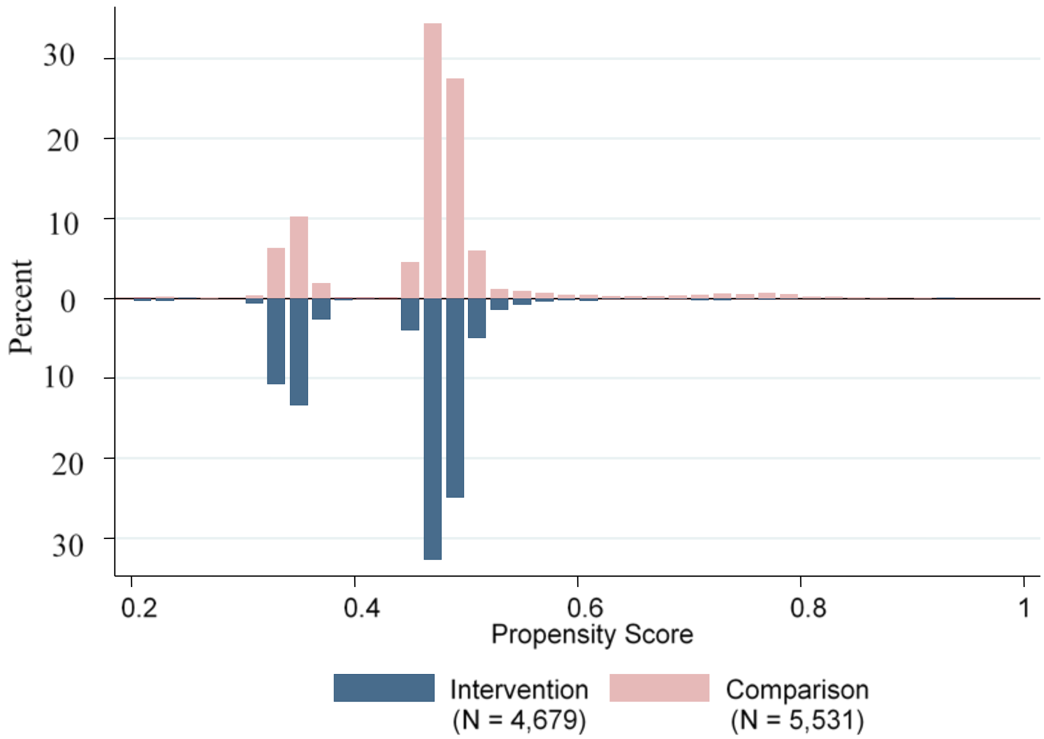
Table A.1. Distribution of propensity scores by intervention status

Intervention status	N	Mean	Standard deviation	Min	10th % ile	25th % ile	50th % ile	75th % ile	90th % ile	Max
Comparison	5,531	0.446	0.077	0.209	0.338	0.358	0.471	0.486	0.500	0.928
Intervention	4,682	0.469	0.086	0.208	0.344	0.461	0.475	0.490	0.512	0.964

Sources: Mathematica Analysis of Tanzania Energy Sector Baseline Household Survey.

There were two intervention households for which their propensity score was greater than the maximum score of the comparison households (that is, the intervention households were off-support), and one intervention household with a propensity score less than the minimum of the comparison households. These three households were dropped from subsequent analysis.[50] Figure A.2 shows substantial overlap between the propensity scores of the 5,531 comparison and remaining 4,679 intervention households, weighted by the nonresponse adjusted sample weight for the intervention group and one for the comparison group. Although there was a great deal of overlap around the two modes, there were fewer comparison households at the right tail of the distribution and somewhat more in the modal group below 0.4.

Figure A.2. Distribution of propensity scores used to construct matched sample weights



b. Matching and weight construction

The propensity score was used to perform the kernel matching and to construct the matched sample weights. Kernel matching is a nonparametric technique that uses the weighted averages of all observations in the comparison group to construct a matched comparison group. Larger weights are assigned to comparison households that are closer to intervention households in terms of propensity score. Thus, for example, the comparison households with propensity scores in the right tail of the distribution, as shown in Figure A.2, will receive larger weights relative to those near the modes. To describe this process, define T to be the set of intervention households and C to be the set of comparison households. Similar to Heckman et al. (1998), each comparison group member i was assigned a matched sample weight using the following formula:

[50] After dropping these three observations, the differences were still statistically significant (t = -14.29)

(2) $$W_i^M = \sum_{j \in T} W_i^{KM}(j),$$

where *j* is the index for intervention households and $W_i^{KM}(j)$ is a weight based on the kernel matching given by

(3) $$W_i^{KM}(j) = \frac{W_j W_i K\left(\hat{q}_j - \hat{q}_i\right)}{\sum_{k \in C} W_k K\left(\hat{q}_j - \hat{q}_k\right)}$$

And K(.) is a symmetric Gaussian kernel function

(4) $$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{u}{h}\right)^2}$$

where *h*, the bandwidth, is positive. The weights for the comparison group households are set to one in equation (3). Following Silverman (1986), we select the optimal bandwidth that minimizes the mean integrated squared error given by

(5) $$h = 0.9 * A * N^{-\frac{1}{5}}$$

where $A = \min(IQR/1.34, \hat{\sigma})$ of the distribution of the propensity scores $\hat{q}$, IQR is the interquartile range of the sample and *N* is the number of households.

Intuitively, when matching to intervention household j, equation (3) assigned a weight $W_i^{KM}(j)$ to comparison *i* that decreased in the difference in propensity scores $\left|\hat{q}_j - \hat{q}_i\right|$ due to the shape of the kernel. Using equation (2), we summed these comparison weights across all intervention households, and the resulting W_i^M matched sample weights were used to estimate baseline differences. Because the kernel matching process did not change the intervention household weights, we defined $W_j^M = W_j$ for each intervention household.

c. Assessing match quality

After we conducted the kernel matching, we found no statistically significant differences between the intervention and comparison households for the individual variables in our model, and all of the variables were jointly insignificant. These results indicate that our model reduced the differences between the two groups along the covariates included in the propensity score model. We ran a series of linear regressions in which each characteristic was regressed on intervention status, first weighted by our initial sample weights W (pre-match) and then by our matched sample weights W_i^M (post-match). The standard errors in each regression were adjusted to account for clustering at the community level. In our pre-match regressions, intervention status was statistically significant at the 0.05 level (two-tailed test) for all variables except:

- Gender of the household head
- Presence of any phone
- House has two rooms
- Between 0–90,000 TZS spent on energy per year
- Between 90,001–480,000 TZS spent on energy per year
- Between 840,000–3,204,000 TZS spent on energy per year
- Non-electric energy made per month, including from small batteries (kWh)

After applying our match weights, no variables were statistically significant at the 0.10 level (two-tailed test). In our joint significance tests, we ran a logistic regression of intervention status on the vector of characteristics included in the propensity score model. We first ran the regression using the initial sample weights and then with the matched sample weights, adjusting the standard errors for clustering. The Wald χ^2 statistic for the pre-match model was 72.91 (df = 18)[51] with $p > \chi^2 = 0.00$ indicating that the variables were jointly significant in predicting intervention status. When our matched sample weights were applied, we fail to reject the hypothesis that the variables are jointly insignificant in predicting intervention status (Wald $\chi^2 = 8.26$, df = 18, $p > \chi^2 = 0.97$). Finally, we conducted a t-test of the propensity score by intervention status, weighted by the newly created matched weight. Prior to matching, the difference between the propensity scores was statistically significant, as discussed above. After matching, however, the differences were statistically insignificant (t = -1.28).

[51] The variable indicating a house with more than two rooms was dropped due to collinearity. As a result, there were 18 degrees of freedom for the chi-squared test rather than 19.

This page has been left blank for double-sided copying.