Evaluation of Tanzania Energy Sector Project: Updated Design Report

March 16, 2011

Duncan Chaplin
Arif Mamun
Thomas Fraker
Kathy Buek
Minki Chatterji
Denzel Hankinson

MATHEMATICA
Policy Research, Inc.

Evaluation of Tanzania Energy Sector Project: Updated Design Report

March 16, 2011

Duncan Chaplin
Arif Mamun
Thomas Fraker
Kathy Buek
Minki Chatterji
Denzel Hankinson

Contract Number:
MCC-05-0192-CFO-TO05

Mathematica Reference Number:
06510.016

Submitted to:
Millennium Challenge Corporation
875 15th St., NW
Washington, DC 20005
Project Officer: Laura Rudert

Submitted by:
Mathematica Policy Research
600 Maryland Avenue, SW
Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763
Project Director: Arif Mamun

MATHEMATICA Policy Research, Inc.

ACKNOWLEDGMENTS

We would like to thank a number of people who provided invaluable advice and guidance on this design report. These include Laura Rudert, Anne Rothbaum Pizer, Jack Molyneaux, Matthew Kavanagh, Karl Fickenscher, and Himesh Dhungel at MCC; Bernard Mchomvu, Ahmed Rashid, Paschal Assey, Salum Ramadhani, Chedaiwe Luhindi, Issac Chanji, Peter Kigadye, Florence Gwang'ombe, and William Christian at MCA-T; Boniface Njombe and Robert Semsella at TANESCO; Ali Maalim and Raymond Noel at ZECO; Keith Treviss at Crown Agents USA; Howard White at the International Initiative for Impact Evaluation (3ie), Douglas Barnes at the World Bank, Jeffrey Grigg at the University of Wisconsin; Lauren Pierce and Brendan Larkin-Connolly of DHInfrastructure; Leonard Sibomana and Zaharani Kalungwa, our consultants in Tanzania; and a large number of other staff at MCC, MCA-T, and the Tanzanian Government. The design report benefited greatly from discussions with our colleagues at Mathematica: Phil Gleason, Nancy Murray, Anu Rangarajan, Frank Potter, Jane Fortsen, Steve Glazerman, and Alison Wellington. We would also like to thank Ebo Dawson-Andoh for excellent research assistance, Carol Soble and Sharon Peters for valuable editorial assistance, and Alfreda Holmes and Donna Dorsey for production support.

CONTENTS

TABLES

FIGURES

ACRONYMS

DID	Difference in Differences
ES	Effect Size
GDP	Gross Domestic Product
GPOBA	Global Partnership for Output-Based Aid
GPS	Global Positioning Unit
ICC	Interclass Correlation Coefficient
ITT	Indicator Tracking Table
IV	Instrumental Variables
MCA-T	Millennium Challenge Account-Tanzania
MCC	Millennium Challenge Corporation
MDI	Minimum Detectable Impacts
MoF	Ministry of Finance
NBS	National Bureau of Statistics
NORAD	Norwegian Agency for Development Cooperation
NRECA	NRECA International
PPΔ	Percentile Point Changes
PSU	Primary Sampling Unit
SIDA	The Swedish International Development Cooperation Agency
T&D	Transmission and Distribution Systems Rehabilitation and Extension
TANESCO	Tanzania Electric Supply Company
ZECO	Zanzibar Electricity Company

I. INTRODUCTION

Access to reliable, high-quality electricity can be a key driver for economic growth (Barnes 1988). In Tanzania, only 2 percent of rural households have access to the government's electricity distribution network, and available power is subject to frequent surges and interruptions in service. With a gross domestic product (GDP) per person of only $510 (World Bank 2009), Tanzania is one of the poorest countries in the world. 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). To address concerns that a lack of electricity may be hindering economic growth in Tanzania and keeping families poor, the government of Tanzania is implementing an energy sector project using funds obtained from the Millennium Challenge Corporation (MCC). This project is intended to increase the supply of reliable and high-quality electricity to people in mainland Tanzania and in Zanzibar.

MCC has contracted with Mathematica Policy Research to conduct a rigorous evaluation of the energy sector project which is being implemented by the Millennium Challenge Account–Tanzania (MCA-T). This report describes Mathematica's evaluation design for the energy sector project based on deliberations with MCC, MCA-T, the Tanzania Electric Supply Company (TANESCO), the Zanzibar Electricity Corporation (ZECO), and other stakeholders over the past two and a half years.

A. Background

MCC is an innovative and independent U.S. foreign aid agency whose primary mission is to reduce poverty through economic growth. It strives to achieve this mission in a number of ways. First, it incentivizes good governance by providing large grants to governments of low-income countries that have demonstrated a strong commitment to sound policy practices. Second, it partners with these governments to help them select and implement effective development programs. Finally, MCC produces rigorous evidence on what works to help inform future poverty alleviation policies and programs.

Tanzania is one of a handful of nations that has been awarded a compact from MCC. At approximately $700 million, the Tanzania compact is the largest awarded by MCC 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 Tanzania, MCA-T is using the MCC compact to fund projects in three sectors: roads, water, and energy. This report presents a design for a rigorous evaluation of the energy sector project. The evaluation will inform MCC, other donor organizations, and researchers as to whether investing in electrification is effective in promoting economic growth and reducing poverty.

MCC is funding three activities under the energy sector project:[1]

[1] MCC is also funding a fourth activity–extending a mini-grid system in the Kigoma region. An evaluation of that activity may be designed later. 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.

- **Transmission and Distribution Systems Rehabilitation and Extension Activity (T&D activity).** This activity involves rehabilitation of existing distribution networks as well as construction of new lines in Tanga, Dodoma, Morogoro, Iringa, Mwanza, and Mbeya—regions identified as being high priority for investment in electricity.

- **Pilot Program of Subsidies for Connections to the T&D Line Extensions (subsidy pilot activity).** 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 activity, to subsidize costs of limited-use connections for households with access to the new T&D activity lines.

- **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.

To evaluate the impacts of the line extensions created through the T&D activity, Mathematica proposes to use a difference in differences comparison group analytic design. To evaluate the impacts of the subsidy pilot activity, Mathematica will use random assignment. For the cable activity, Mathematica proposes two distinct pre-post analyses: The primary one will be based on data on electricity use, reliability, and quality from ZECO. This will be supplemented by a case study of a small set of larger hotels on Ungjua Island.

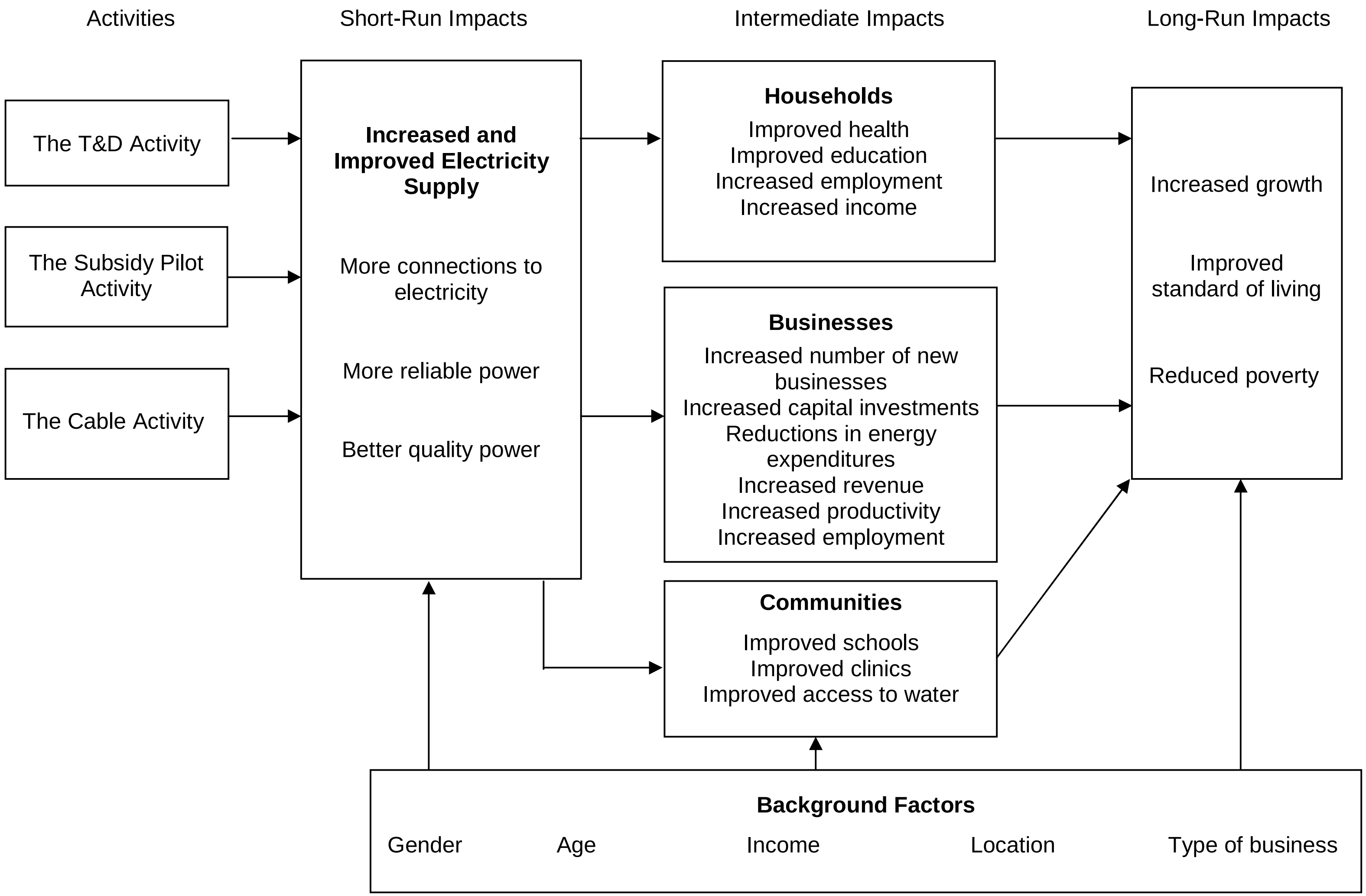
B. Conceptual Framework

Mathematica has developed a conceptual framework, presented in Figure I.1, to guide our approach to the evaluation of the three activities under the energy sector project. The boxes on the far left of the figure show the three energy sector activities. The box on the far right shows the ultimate objectives of the activities—increased growth, an improved standard of living, and poverty reduction. The activities are designed to achieve these objectives through their effects on electricity supply, 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 the electricity supply in three ways, as shown in the box in the second column of the conceptual framework. First, by increasing the reach of the distribution networks, the T&D and subsidy pilot activities can increase the number of households, businesses, and community organizations (such as schools, health facilities, and water utilities) with electrical connections. Second, the activities are designed to reduce the extent of service interruptions or outages, referred to as the "reliability" of electricity supply. Third, the activities may reduce variations in voltage magnitude or harmonic distortions, referred to as the "quality" of electricity supply. Poor-quality electricity can lead to equipment damage at the electric utilities and in homes and businesses.

Increased and improved electricity supply can have important intermediate impacts on households, businesses, and communities, as shown 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 and 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 they wish to

Figure I.1. Conceptual Framework for the Energy Sector Project



take advantage of via radio and television programming and cell phone communications. Electricity can improve health outcomes if it enables households to reduce use of certain types of fuel that can be particularly bad for the health, such as charcoal for cooking. 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 use electricity for refrigeration and certain types of medical equipment), and water utilities (which can use electricity for pumps and cleaning equipment). For all of these types of uses, the additional grid electricity funded by MCC can be far less expensive than electricity produced by the small generators commonly used by large hotels and some businesses operating far from current electric grids.

The box at the bottom of the framework shows background factors that may affect the short-run, intermediate, and long-run 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 benefit 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.

In the remainder of this report, we describe how we are designing evaluations for each of the three activities (Chapter II), our power calculations (Chapter III), and our data sources (Chapter IV). In Chapter V, we explain how we will address our key research questions within the evaluations. In Chapter VI, we describe a timeline for various tasks related to the evaluation. In the concluding chapter, we discuss potential challenges, proposed responses, and next steps. In Appendix A, we briefly discuss alternative designs considered for the T&D evaluation. In Appendix B, we provide a timeline for our study.

II. TECHNICAL APPROACH

Our goal is to estimate the impacts of the energy sector activities discussed above and to use the results to answer a set of MCC-prescribed research questions. Achieving this goal will be challenging because observed outcomes can be influenced by factors other than these activities. Our technical approach is therefore designed to isolate the effects of electrification activities from other potentially confounding factors. To that end, we will employ the most rigorous evaluation methods possible without compromising the MCA-T goals and implementation plans. In principle, an ideal evaluation method would compare outcomes for households, businesses, and communities that received improved electrical supply (new lines, subsidies, or a new cable) with outcomes for the same groups if they had not received the improved supply. In practice, however, once customers have received improved electrical supply, it is impossible to observe what would have happened if they had not received it. The effects of the improved electrical supply may be approximated only by comparing outcomes for the households and business that received the intervention with outcomes for similar groups that did not receive the intervention. These groups, which represent the counterfactual, could be of the three following types: (1) the same households, businesses, and communities before the intervention was implemented (the pre-post method), (2) groups similar to those that received the intervention in terms of observable characteristics (the difference in differences comparison group method), or (3) groups that did not receive the intervention but are similar to those that did receive it in terms of both observable and unobservable characteristics (the random assignment method). We propose to use all three methods, with the choice of method depending on the activity to be evaluated.

MCC and MCA-T funded the Tanzania electrification activities based on careful study of their likely benefits and costs, necessitating extra care in selecting groups to represent the counterfactual. To maximize the projected benefits of the activities, electrification is being implemented in areas with relatively high concentrations of households. To minimize the costs of the activities, most of them are being implemented close to existing power lines. Other factors also may have influenced the selection of sites for the electrification activities, such as the presence of government offices. Consequently, households and businesses located in the activity sites are likely to be different in observable and unobservable ways from those located in other areas. This means that a simple comparison of outcomes between households and businesses that receive the interventions and those that do not receive the interventions could easily yield biased impact estimates due to confounding of the effects of the intervention with initial differences between the groups. A rigorous evaluation with well-designed groups to represent the counterfactual, such as those described above, will minimize the likelihood of obtaining impact estimates biased by the initial differences between those who receive the intervention and those who do not.

Random assignment is the most rigorous way to estimate impacts of an intervention. Researchers can use this method to create two groups that are virtually identical on average except that one group (the treatment group) was exposed to the intervention while the other (the control group) was not. As a result, any differences in outcomes between the two groups that are too large to be attributable to chance may be attributed to the effects of the intervention. However, random assignment is not always feasible, such as in the evaluations of the T&D activity and the cable activity. In evaluating these activities, we will strive to create groups that represent the counterfactual that most closely approximates the benefits of random assignment, thus avoiding potential selection bias in the impact estimates. We will also strive to minimize the burden that our evaluation designs impose on MCA-T and their subcontractors who are implementing the electrification activities. In Table II.1, we summarize our proposed technical approach to the evaluation of each activity. It

Table II.1. Summary of Technical Approach to the Evaluation of Each Energy Sector Activity

<table>
<tr><th>Activity</th><th>Evaluation Methodology</th><th>Intervention/Treatment Group</th><th>Counterfactual: Comparison/Control Group</th><th>Key Outcomes</th></tr>
<tr><td>T&D</td><td>Difference in differences (DID) method, which compares changes in outcomes over time between T&D intervention and comparison areas</td><td>Households, businesses, and communities in areas that received line extensions</td><td>Households, businesses, and communities in areas that do not receive new line extensions</td><td rowspan="4">Access to, reliability and quality of electrical power[a]
Household income and expenditures
Business energy expenditures
Employment
Health outcomes
Child schooling attainment (or at least intensity of study)
Distribution of time and resources within the household by gender</td></tr>
<tr><td>Subsidy Pilot</td><td>Random assignment of areas either to a treatment or control group; compare outcomes between these two groups at follow-up</td><td>Households in areas that received connection subsidy offers</td><td>Households in areas that do not receive connection subsidy offers</td></tr>
<tr><td rowspan="2">Cable</td><td rowspan="2">Pre-post approach comparing outcomes from before and after implementation of the cable activity</td><td>Months after activity is completed</td><td>Months before activity begins</td></tr>
<tr><td>Hotels on Unguja Island after activity is completed</td><td>Hotels on Unguja Island before activity is completed</td></tr>
</table>

[a]Quality of electricity refers to voltage fluctuations while reliability refers to interruptions in the power supply.

identifies the evaluation methodology that we will use to control for the counterfactual, the group or groups that will receive the intervention or represent the counterfactual, and key outcome measures.

Our evaluation is designed to answer the following research questions regarding the electrification activities:

- **Impacts on implementation measures:** What is the impact of the project on access to electricity, energy quality, and energy reliability?
- **Impacts on outcomes:** Does access to electricity lead to (1) increased household income and better health and education outcomes; (2) increased business activity, including creation of new firms, capital investments, and greater levels of employment; or (3) improved community outcomes related to schools, hospitals, or water supply and, if impacts are detected, what are the magnitudes of those impacts?
- **Benefit-cost analyses:** Is the project warranted based on re-estimation of benefit-cost analyses?
- **Analyzing the benefit of a rigorous evaluation:** Does a rigorous evaluation design yield the same impact estimates as a simple before-and-after comparison group design?
- **Subgroup analyses:** Do the impacts vary by gender, age, and income?
- **Unintended consequences:** Are there unintended impacts of the program (positive or negative)?
- **Lessons learned:** What are the implications of the evaluation findings for replication and long-term policymaking?

We will describe our evaluation results in two reports—an interim report that will be completed before the end of the compact and a final report that will be completed later. The interim report will cover short-run findings based on administrative data on electricity use, reliability, and quality as well results from analysis of the qualitative data. The final report will cover longer term impacts on these same outcomes based on administrative data, and also a rich set of additional outcomes based on survey data related to poverty, economic development, and well-being. To further illustrate our findings we will include case studies of businesses in the Tanga region, and a small set of larger hotels in Zanzibar in the final report. Throughout we will carefully evaluate and disaggregate results by gender given the importance of this issue for MCC.

A. Evaluation Design, Sampling, and Analysis Plans for the T&D Activity

We propose to use a difference in differences (DID) method 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[2] communities (i.e., those that will receive the line extensions) with changes in outcomes for comparison communities. The comparison communities will be chosen to be similar to the intervention communities based on various household and community characteristics, such as whether the area is already connected to the power grid, the types of

[2] 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 subsidies through the subsidy pilot activity. We refer to that subset as the "treatment group."

materials the homes are made of, and the types of energy used for cooking and lighting. 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.

The intervention and comparison communities will be selected from a set of primary sampling units (PSU) similar to those used by the National Bureau of Statistics (NBS) in Tanzania. In particular, we will use subvillages (*vitongoji*) in rural areas and neighborhoods (*mtaa*) in urban areas.[3] These subvillages and neighborhoods have elected leaders and boundaries defined by the government. The subvillages and neighborhoods used in our study will be selected in six steps. First, we will finalize the current list of areas that are likely to receive line extensions.[4] Second, we will develop a list of villages and urban neighborhoods in these areas. Third, we will randomly select 182 of those villages and neighborhoods for the study. The number should be sufficient to obtain the desired level of precision, as discussed below in our section on statistical power. Fourth, in each rural intervention village, we will select a subvillage expected to have a high fraction of households with access to the new lines. Fifth, we will select comparison villages and neighborhoods to be similar to those in the intervention group. Sixth, in each comparison village, we will select a subvillage that is matched to the population rank of the corresponding intervention subvillage.

The fourth step (selecting a subvillage expected to have a high fraction of households with access to electricity) is needed because most villages in Tanzania consist of a number of smaller subvillages, and we expect that, in many cases, not all of the subvillages will have substantial access to the new lines.[5] Hence, in rural areas, we recommend that the evaluation focus on the subvillages with substantial access. This step is not needed in urban areas as a *mtaa* is the smallest administrative unit in urban areas, and we expect that, in urban neighborhoods that receive line extensions, almost all households will have access. Because we are selecting intervention subvillages that are expected to have high fraction of households with access, our results will not generalize to households in communities where a small fraction of households have access to electricity. However, it is conceivable that providing access to electricity to all households will be the long run policy objective. Consequently, estimating impacts for subvillages with greater fraction of households having access to electricity would be more policy relevant than estimating impacts for subvillages with a small fraction of households having access. We discuss the generalizability of our results in greater detail in section D below.

The fifth step, selection of comparison subvillages and urban neighborhoods, will proceed as follows. For each of the 182 intervention communities, we will identify a matched comparison community by using propensity score matching, a statistical method of matching based on several criteria (Rosenbaum and Rubin 1983). The matching will be implemented in two stages. In the first stage, we apply a nearest-neighbor matching with replacement method and use existing census and global positioning system (GPS) data from NBS as well as data from TANESCO to identify 546 potential comparison communities, which is three times the number of intervention communities. NRECA International (NRECA), the firm contracted to carry out various surveys for this

[3] NBS also uses subvillages and neighborhoods, but, with the boundaries changing over time, it is unlikely that the set we use will correspond exactly with the set NBS uses for any particular survey.

[4] We have 337 areas on our current list, divided into 182 subprojects. Subprojects are units used by the engineering firms building the lines.

[5] By access, we mean being within a certain distance from a pole on the new low-voltage line. Households or businesses within this distance are eligible for a connection at a basic rate. Entities farther away must pay for additional poles. Currently, the distance is 30 meters, but TANESCO is considering an increase to 60 meters.

evaluation, will then implement a community survey in the 182 selected intervention communities and 546 potential comparison communities for a total of 728 communities. In the second stage of propensity score matching, we will apply a matching without replacement method[6] and use data collected in the community survey to identify one matched comparison community for each intervention community. This means that we will end up with 182 intervention communities and 182 comparison communities in which we will conduct the baseline and follow-up household surveys.

We will apply the following strategy to implement the last step, selection of a subvillage in each comparison village. All subvillages in each intervention and comparison villages will be ranked by population size (measured by number of households). We will note the rank of the subvillage selected in each intervention village, and then in the corresponding matched comparison village we will select the subvillage that has the same rank order. Thus, the subvillages in a pair of intervention and comparison villages will be matched on their relative size within the village.

In addition to the community survey mentioned above, we will conduct household and enterprise surveys in the T&D study. For the baseline household survey, we will randomly sample households from each intervention and comparison community. We do not expect household that are already connected to the grid or close to an existing line will connect to the new lines. Consequently, they will be excluded from the survey.[7] We will sample from the remaining households with equal probability within each subvillage or urban neighborhood, which means that we will interview more households in the larger subvillages and neighborhoods. For the enterprise survey, we will use sampling methods similar to those for the household survey, except as noted below.

In the T&D study, we will collect enterprise data in both the household and enterprise surveys. In the household survey, we plan to ask each household about any enterprises it owns. While we expect to see few enterprises identified in our household survey in any one community, aggregation of the data across all households in our sample (over 6,000) should produce information on a reasonably large number of businesses. In addition, when comparing the baseline and follow-up survey results, we should be able to determine any increase in business activity. The enterprise survey will complement the household survey by providing more detailed information on larger businesses. However, to keep this part of our study cost-effective, we are limiting it to a case study of only a small number of businesses and only to the Tanga region.

We will use data from the household, community, and enterprise surveys to estimate impacts of the T&D activity by using the DID method. To estimate impacts using DID, we will estimate equations of the following form:

$$Y_f = \alpha + \beta_1'I + \beta_2'Y_b + \beta_3'X + e \qquad (1)$$

where Y_f = outcome for a household, business, or community from the follow-up survey; I = dummy variable for being in an intervention community; Y_b = outcome at baseline; X = other

[6] "Without replacement" means that each comparison community will be matched to just one intervention community.

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

characteristics at baseline; e = an error term; and β_n are coefficients describing the relationships between the outcome and each of the right-hand–side variables (with β_1 representing the impact of being in a community with a new line extension). In estimating this equation and the equations presented below, we will use linear regression when the outcome is a continuous measure and an appropriate functional form (e.g., logistic or probit) when the outcome is discrete.

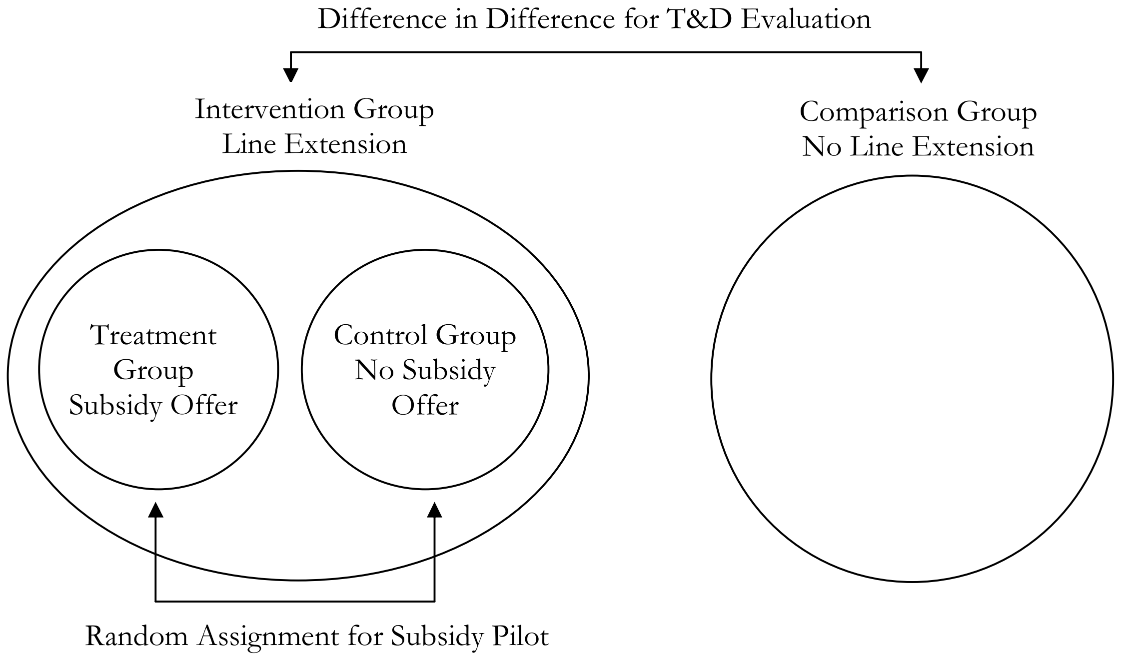
B. Evaluation Design, Sampling, and Analysis Plans for the Subsidy Pilot Activity

Electric line extensions have the potential to yield significant benefits for poor households in Tanzania. Unfortunately, many such households cannot afford to connect to the lines. For this reason, key stakeholders are interested in learning more about the impacts on poor people of subsidized access to electricity. Indeed, during a visit to Tanzania in 2009, Mathematica staff met with representatives from several funding organizations, such as the World Bank, the Norwegian Agency for Development Cooperation (NORAD), and the Swedish International Development Cooperation Agency (SIDA). These funders expressed strong interest in learning more about the impacts of subsidies for electrical connections. In a recent pilot project in Tanzana, SIDA subsidized 85 percent of the costs of electrical connections for almost 1,000 customers (Sweco, 2008). We suspect that other funders, such as the Global Partnership for Output-Based Aid (GPOBA), might also be willing to subsidize connection costs if presented with strong evidence that such subsidies improve well-being. Indeed, GPOBA has already provided such subsidies in Armenia.

In response to this interest, MCC is funding a study of the subsidy pilot activity in conjunction with the line extension study. This activity will subsidize the costs of connections to the extended power lines, making it more likely that those households will connect and thereby helping to ensure the availability of better information about the impacts of electricity on the poor. In particular, we will estimate impacts of the subsidy offers on connection rates and on the same outcomes covered in the T&D study. This information will help TANESCO, policymakers, and donors as they consider subsidy arrangements in the future.

The subsidy pilot activity is being implemented only in the communities covered by the T&D line extensions. It will cover all six of the regions in the T&D study. Therefore, the evaluation of the subsidy pilot is closely related to the evaluation of the T&D activity, as illustrated in Figure II.1. Both the treatment and control groups for the subsidy pilot evaluation will be selected from among the intervention communities for the T&D evaluation. We will randomly assign up to half of the T&D intervention communities to a treatment group to be offered subsidies for connections to the line extensions. The remaining intervention communities will constitute a control group that will not be offered subsidies. An even split of the T&D evaluation communities between the treatment and control groups for the subsidy pilot study would maximize the statistical power of the subsidy analysis. However, current uncertainty regarding available funding and the proportion of treatment households per community that will take up the subsidies could mean that we will not be able to assign that many communities to the treatment group. Statistical power would not be greatly reduced if we end up with a somewhat smaller fraction of communities in the treatment group, provided the fraction is not too far from half, as discussed in Chapter III.

Figure II.1. T&D and Subsidy Pilot Evaluations



In order to implement the subsidy pilot effectively the evaluator will be hiring a communications firm that will be charged with contacting the community leaders and letting them know about the study, organizing public events for random assignment, inform eligible households about the subsidy offer and how they can apply for these subsidies. Hankinson et al (2011) provides a detailed description of how the subsidy pilot will be implemented.

The design for the evaluation of the subsidy pilot activity 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 subsidies and the outreach work of the communication firm for the part of our T&D intervention group to be offered subsidies. Second, we will be able to estimate impacts of the line extensions without subsidies by excluding the treatment communities for the subsidy pilot activity when estimating the impacts of the T&D activity, though the resulting estimates will be somewhat less precise than our main results.

One goal of the subsidy pilot impact evaluation is to understand how offering subsidies for electricity connections affects poor households that cannot afford to connect on their own. This goal will be difficult to achieve because some households in the study will accept subsidies but would have connected on their own even without the subsidies. They will still benefit from the subsidy, but not because it changes their use of electricity. Rather, it will be equivalent to an offer of extra income for those households.

We propose to address this issue in three ways. First, we will limit the subsidy offers to owner-occupied households, largely eliminating the possibility that a developer will buy up several homes, get them all electrified, and then sell them at a substantial profit. It will also reduce the likelihood that a poor household is forced to move because its rent increases after electrification.

Second, we propose to reduce further the number of households that would connect on their own in the subsidy-treatment group by restricting the subsidies to households eligible to use ready boards. Ready boards are a substitute for regular internal wiring and may be used by households that plan to use only a few light bulbs or small appliances. By eliminating the need for internal wiring, ready boards can save a household hundreds of dollars. TANESCO generally does not approve the use of ready boards in houses with more than two rooms because of safety concerns. Consequently, we expect that many households that can afford to connect on their own and want electricity for more than a few light bulbs will not use the ready-board option. While we will not be able to replicate the ready-board eligibility criteria in the control group, we will be able to exclude large households from our study sample in both the treatment and control groups, helping us focus our evaluation on households that are less likely to be able to connect without the subsidy.[8] Restricting the subsidies to households eligible for ready boards will also enable us to provide subsidies to a larger number of households able to connect only with subsidies, thereby increasing the statistical power of our study.

Third, as discussed in the following paragraphs, we propose to use our survey data to estimate the extent to which our results are affected by households that would connect without the subsidy. While limiting our subsidy pilot evaluation to households that are owner-occupied and ready board-eligible should help, some households in our sample would accept ready-board subsidies even though they would have connected without the subsidy. To deal with this complication, we will use survey data collected for the study to estimate this fraction. We will then use that estimate to help us interpret the principal statistical results from the study. Suppose, for example, that we estimate that 5 percent of households offered subsidies would have accessed electricity without the subsidies. For these households, the subsidy offer would be the same as being offered $300 in cash, thus increasing their ability to purchase and consume goods and services. Since this subgroup is not relevant for policy purposes, we will interpret our results with such households in mind.

When analyzing the impact of the subsidies on total household consumption (excluding the electricity connections themselves), it may be possible to adjust our estimates to remove this subgroup. Continuing with the example above, to estimate the impact on consumption for households that would not connect without a ready-board subsidy, we could subtract $15 (5 percent of $300) from our overall estimate of the impact of the subsidy offer across all households and then divide the result by 1 minus the fraction that could connect on its own. If the original estimate across all households was $34, then the updated estimate would be $20 = ($34-$5)/0.95.

To see why this works, we note that the overall effect may be written as a weighted average of the impacts for those who accept subsidies even though they would have connect on their own and the rest of our sample. Thus,

(1a) $$I_o = P_c * I_c + P_r * I_r$$

where I_o = overall impact on consumption; c = group that gets a ready board if in the treatment group that would connect on their own without a subsidy; r = remainder of households (the ones

[8] This also means that we are unlikely to be covering larger home-based businesses. We will not be subsidizing connections to businesses that are not in the home, but are likely to be covering at least a few home-based businesses in smaller homes.

that are of interest for the evaluation); P_k = fraction of households in group k where k=c (for connect on own) or r (for remainder); and I_k = impact for households in group k.

Now if we solve for I_r, we get:

$$I_r = (I_o - P_c * I_c) / P_r \quad \text{(1b)}$$

To estimate I_r using equation (1b) requires estimates of the other quantities. I_o is the estimated impact overall that we obtain by using our standard methods. P_r equals one minus P_c. P_c can be estimated as follows: Households that use the ready-board subsidies when in the treatment group--even though they would connect without such subsidy if they were in the control group--can come from two subgroups. The first is the subgroup that would connect with a ready board if in the control group. The second is the subgroup that would connect with regular wiring if in the control group. We can estimate the size of the first subgroup by using the fraction of households in the control group that gets ready boards. We expect this group to be small because ready boards are not yet common in Tanzania. To estimate the size of the second group, we start with the fraction of the control group connected with regular wiring (which includes some who would use subsidies for ready boards if offered) and then subtract the fraction of households getting a regular connection in the treatment group (which excludes the group that takes ready-board subsidies). This second group could be moderately large if some small owner-occupied households decide to get ready boards for now even though they are planning to switch to a regular connection later. This seems plausible since there is no cost to changing from a ready-board connection to a regular connection. Hence, for these households the subsidy offer can be used to cover their regular connection costs as long as they are willing to postpone getting the regular connection. They do still have to qualify, however, for the subsidy offer, which is based on being qualified to use a ready board (that is, have a small house) and being owner-occupied.

Some households that cannot afford electricity may be willing to accept a subsidized ready-board connection, meaning that they could not benefit from their connection to an extended power line. However, we believe that the number of such households is small because TANESCO offers a very low-cost option (less than $20 per year) for sufficient electricity to operate several light bulbs. To help ensure that we do not subsidize households that could not afford this small charge, we plan to require households to pay something to get connected. For example, we might require them to pre-pay for some electricity, perhaps around $10.

Saving $10 may require significant time for some households. In addition, it takes time to apply to TANESCO for a connection, with or without a ready board. Consequently, we plan to give households at least one month to connect after the offer of a subsidized connection. Since we cannot identify households in the comparison group that could not afford this payment, we will include in our analyses all owner-occupied, ready board-eligible households in both the treatment and comparison groups.

To estimate impacts of the effect of the offer of the subsidy, we will estimate equations of the following form:

$$Y_f = \alpha + \beta_1{'}T + \beta_2{'}Y_b + \beta_3{'}X + e \quad \text{(2)}$$

where Y_f = outcome for a household at follow-up survey; T = dummy variable for being in a treatment community (one where eligible households were offered subsidies for connections); Y_b = outcome at baseline; X = other characteristics at baseline; e = an error term; and β_n are coefficients describing the relationships between the outcome and each of the right-hand–side variables (with β_1 representing the impact of the subsidy offer).

The primary criterion we will use to determine eligibility for using ready boards is the size of the dwelling. The information we have been given suggests that TANESCO gives its field engineers some discretion when determining which households are eligible for ready boards. The goal appears to be minimizing risk of electrical fire that might occur if households try to use the ready boards for too many electric appliances. Thus, we have been told that in general TANESCO limits ready board use to homes with only one or two rooms. Since there is some uncertainty about exactly how a TANESCO engineer would decide on a given house, our estimates of which households would be eligible for using ready boards will be approximate.

Our goal for the subsidy pilot evaluation is to estimate impacts of the offer of a subsidy. As discussed above, we cannot do that exactly because we cannot precisely replicate the method TANESCO uses to identify households that can use ready boards. We can, however, limit our analyses to households that appear to be subsidy-eligible based on our survey data. We expect to miss a few eligibles (houses with more than two rooms that appear ineligible based on the survey data but were considered acceptable by TANESCO, perhaps because the TANESCO engineer decided that the house in question was unlikely to use much electricity), and we expect to include a few ineligibles (houses with one or two rooms that appear to qualify based on survey data but were turned down by TANESCO, perhaps because of concerns about the building materials).[9] This means that our results will not generalize to the former group and that our impact estimates will be somewhat diluted because of the latter group. We will keep these limitations in mind when interpreting our results. In addition, we will ask TANESCO to provide us with estimated ready-board eligibility by household for a randomly chosen set of households and will use those data to estimate the fraction of eligibles we are excluding and the fraction of ineligibles we are including when we rely on the survey data to determine eligibility. We will use those fractions when interpreting our results

C. Evaluation Design and Analysis Plans for the Cable Activity

The cable activity involves laying an underwater cable to connect Unguja Island (the principal island in Zanzibar) with the mainland, thus improving the reliability and quality of electricity service in Zanzibar. The improvement is likely to occur at about the same time for all customers, thus rendering the creation of a comparison group in Zanzibar impossible. Consequently, we will use a pre-post design to analyze the monthly administrative data from the indicator tracking table (ITT) that ZECO compiles for MCA-T on electricity use, reliability, and quality for all of Unguja before and after the laying of the cable. To obtain the most precise estimates possible, we will use all months of the ITT data that are available when estimating impacts on use, reliability, and quality, except for data during the blackout months, by which we mean power outages that lasted more than one day.[10] Since there were only two blackouts in Zanzibar caused by the cable in recent decades, we

[9] We explored the possibility of using building materials to limit our sample but were told that the rules are not clear for those. For example thatched roofs are not allowed in some situations but are allowed in others.

[10] Because of power rationing ("load shedding") and faults on the cable, there are intermittent power outages in Unguja Island, but these interruptions usually last less than a day. During the two blackout periods, the Island

(continued)

do not expect to have enough statistical power to say anything conclusive about the likelihood that the cable reduces the incidence of blackouts. We will ask for opinions from engineers about the likely impact of the cable on the possibility of future blackouts and include a summary of those opinions in our reports.

We also propose to conduct a case study of the hotel industry in Unguja Island. The hotel case study will consist of two components. First, we will use a pre-post evaluation design to estimate the impacts of the cable activity on key outcomes such as electricity use, reliability, and quality. Baseline data on these outcomes have already been collected from 30 hotels on Unguja Island. We will collect data on these same outcomes from the same hotels after placement of the cable, allowing for a pre-post comparison of the outcomes of the hotels affected by the cable activity. To focus the study on the hotels likely to have the largest impact on the economy of Zanzibar, the 30 hotels were randomly selected from among the 45 largest hotels on Unguja Island. We will collect the follow-up data one year after the cable is completed to allow the hotels to take advantage of their improved access to power. The second component of our hotel study is a description of what hotels reported about the impacts of the two recent blackouts on their business activities. Together these components will help us to develop a richer understanding of the impacts of the cable activity.

The pre-post design cannot definitively estimate causal impacts of the cable activity because it cannot distinguish between changes in the outcome measures that are attributable to the cable and those that may be attributable to other simultaneous interventions or time trends. Changes that might occur simultaneously with the installation of the new cable include increased (or decreased) investment in industries on Unguja Island, changes in demand for the products and services originating on the island, both of which could impact electricity use, or changes in the amount of electricity provided to Zanzibar from the mainland.

To estimate impacts using the pre-post method, we will estimate an equation similar to those for the DID (for the T&D activity), but without a comparison group:

(3) $$Y_t = \alpha + \beta_1\text{'Post} + \beta_2\text{'X} + e_t$$

where Y_t = outcome (pre or post); post = dummy variable for an outcome measured after the cable is completed; X = other characteristics at baseline; e_t = an error term; and β_n are coefficients describing the relationships between the outcome and each of the right-hand–side variables (with β_1 representing the impact of the cable activity).

We may also explore controlling for any time trends in use, reliability, and quality that appear before the introduction of the cable, using various functional forms such as a linear control variable for time or various quadratics (time squared and time cubed). If data are available on the degree to which the mainland limited electricity availability to Zanzibar through rationing then we will control for that as well.

(continued)
experienced power outages for several weeks. The first of the two lasted from May to June 2008 ("the 2008 blackout") and the second from December 2009 to March 2010 ("the 2009–2010 blackout").

D. Internal and External Validity

The evaluation of the Tanzania energy sector project is designed to provide rich information on key issues while cost-effectively balancing internal validity (that is, unbiased estimation of impacts on the study sample) and external validity (that is, the ability to generalize the findings to a larger population). The DID method used to evaluate the T&D activity and the pre-post method used to evaluate the cable activity will provide optimal external validity. Both methods cover the populations that are most likely to directly benefit from the line extension and cable activities in communities that are most affected. The random assignment method of evaluating the subsidy pilot will provide maximum internal validity, producing unbiased estimates of impacts of the subsidy on the study participants. Furthermore, the findings from this study will generalize to the relatively low-income households located near the newly extended electricity lines. In contrast, the enterprise and hotel surveys and case studies will provide less internal and external validity than the other methodologies because of the relatively small numbers of enterprises and hotels in the sample. However, these studies will capture key outcomes that could not easily be covered using the other methodologies.

Our evaluation of the T&D activity will cover households *directly* affected by the newly extended lines in the communities that are most affected. It will not cover households affected *indirectly* or communities that are less affected. Covering households affected indirectly would be impractical because people almost anywhere in Tanzania might be indirectly affected by the line extensions if, for example, they purchased goods produced in those areas or used services such as hospitals, schools, and water supplies in those areas. We will sample households that might be directly affected by the T&D activity, focusing on communities where a large fraction of households are expected to have access to the new lines. We decided to exclude the communities where only a small fraction of households will likely have access to the new lines because including them would unambiguously reduce expected impacts on connection rates and would likely reduce expected impacts on other outcomes. This in turn would reduce our ability to estimate critical parameters needed to calculate the realized economic rates of return (ERRs). Even in the more intensively served communities—where a large fraction of households will likely have access to the new lines—we will exclude households that are already connected to or are living within 30 meters of existing electricity lines because we do not expect the line extensions to affect their connection decisions. Thus, with a high degree of confidence, we will be able to generalize the findings from this study to the full population of households directly affected by the T&D activity in the more intensively served communities. Because in the long-run all communities are expected to be served intensively, the degree to which we can generalize the findings from the line extension study is highly policy relevant. It would therefore be appropriate to use these findings to calculate the realized ERR of the line extension activity.

The sample of households for the subsidy pilot study will be drawn from the intervention group for the line extension study. Our ability to generalize the findings from this study will therefore be similar to that of the line extension study. However, it will be different in one key respect—the subsidy pilot study will focus on ready-board-eligible, owner-occupied households located near the extended lines. We expect these households to be relatively low income (among those using electricity) and most likely to benefit from the subsidy. Thus, the subsidy pilot component of the evaluation will provide results more relevant to low-income households than will the broader DID analysis of the line extensions.

The enterprise survey will provide results relevant to larger enterprises in Tanga, and the hotel survey will provide results relevant to larger hotels on the Unguja Island of Zanzibar. We will not be able to generalize these results to smaller enterprises and hotels, but we can use them as illustrative case studies to help us address several of the research questions considered for this evaluation.

III. 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 proposed for the T&D household and subsidy pilot components of our evaluation are large and should provide sufficient power to detect impacts of policy-relevant magnitudes. The sample size for the cable study of electricity use will be small but we expect large impacts there. The sample sizes for the T&D enterprise and cable hotel surveys are small, so those results will be more illustrative than the others, but still valuable as case studies.

The data for the impact analysis will be at the household, community, and enterprise levels. We will account for clustering by community to ensure correct measurement of precision of estimated impacts. Clustering occurs because residents of the same subvillage or urban neighborhood are likely to face similar, unobserved (by us, the evaluators) random shocks that affect the outcomes, such as adverse weather conditions. This results in greater correlation of outcomes among households or enterprises in the same communities than can be explained by the variables in our models (or than would be the case without clustering). Thus, the correlation between outcomes for households within the same communities must be factored in when determining the sample size needed to estimate effects of a given size. For the evaluations of the T&D and subsidy pilot activities, we estimated intra-class correlations (ICC)—which measure the extent to which households in the same community have similar outcomes—based on data on income and assets from the 2001 Household Budget Survey of Tanzania. We assumed that the same correlations hold for businesses.

Our analysis will also address the statistical problem of "multiple comparisons." This problem arises when we estimate impacts for a large number of outcomes such that at least a few of the estimates likely will be statistically significant by chance, even if no true impact occurred. We will approach the multiple comparisons problem pragmatically—that is, by specifying, a priori, a small number of domains (such as income, education, health) in which we expected to see impacts and identifying the key outcomes in each domain. We will then adjust for multiple comparisons for these key outcomes within each domain (Schochet 2009).

The minimum detectable impact (MDI) is the smallest true impact that can be detected with a given level of power. In Table III.1, we display MDIs by activity and present MDIs for the following two scenarios: (1) assuming no clustering in the data and (2) assuming some clustering, specifically that 13 percent of the variance in outcomes is across rather than within communities. For the subsidy pilot, we assume that only about 70 percent of households are eligible for the subsidies (that is, households are eligible for a ready board and are owner-occupied). We present each MDI both in effect size (ES) units and as percentile point changes (PPΔ). Effect sizes represent impacts measured relative to the size of the standard deviation of the variable, so that an effect size of 0.10 represents a change in a measure of 0.10 of a standard deviation. Percentile point changes estimate how far the mean moves expressed as percentiles of the original distribution. Thus, the 2.8 in the first row means that we should be able to detect the impact of the T&D activity on household income if the intervention changes income from the median (the 50th percentile) to the 52.8th percentile of the comparison group distribution (50 plus 2.8).

Table III.1. Minimum Detectable Impacts (MDI) for Energy Sector Evaluations

				Assuming No Clustering		Assuming 13% Clustering	
Activity	Units	PSUs	N	ES	PPΔ	ES	PPΔ
T&D	Households	364	11,648	0.071	2.8	0.100	4.0
	Businesses	16	64	0.957	33.1	1.175	38.0
	Communities	NA	364	0.401	15.6	NA	NA
Subsidy	Households	182	4,077[a]	0.120	4.8	0.165	6.6
Cable	Months	NA	72	0.660	24.5	NA	NA
	Hotels	NA	30	1.397	41.9	NA	NA

Note: NA =not applicable; ES =effect size; PPΔ =percentile point changes. The estimates with clustering assume that 13 percent of variation in the outcome is at the PSU level. Effect sizes are in standard deviation units. To construct our power calculations, we assumed a confidence level of 95 percent, two-tailed tests, and 80 percent power. For ease of presentation, we estimated our power calculations for one outcome—income. "Months" refers to the ITT data on electricity consumption, reliability, and quality that are available on a monthly basis from ZECO.

[a] This sample size accounts for the fact that we will exclude from the subsidy pilot evaluation all households that do not appear to be eligible for subsidies.

The MDIs for the T&D and subsidy pilot household outcomes range from 2.8 to 6.6 in percentage point units. We recommend using sample sizes that allow for estimating effects of this size in order to account for the possibility that many households will not have access to the lines and that many with access will not connect. Let us suppose that using electricity increases income from the median to the 60th percentile during the course of our study (a fairly large shift). This represents a 10 percentage point shift. If only half the households with access use electricity, we would expect to see only a 5 percentage point shift in our sample. Moreover, if about a third of the households in our sample end up not being close to the new lines, then we would expect to see only a 3.3 percentage point shift in income, which is between the MDI estimates for the evaluation of the T&D activity with and without clustering as shown in the first row of Table III.1.

A similar result holds for the evaluation of the subsidy pilot activity. In Table III.1, we have adjusted the sample size of households in the subsidy pilot to account for the fact that we plan to exclude from our survey sample households that appear to be ineligible for a subsidy. This is expected to reduce our sample size by about 30 percent. We do not expect to be able to perfectly identify eligibility. Hence, we assume that only about 90 percent of the remaining sample are eligible. Suppose, also, that about 80 percent of the treatment group connects to electricity compared to only 20 percent of the control group—thereby implying a 60 percentage point increase in the connection rate induced by the subsidy and that we continue to assume that actually using electricity increases income by 10 percentage points. This would imply that we would expect to see being in the treatment group increase income by 5.4 percentage points (10 times 0.9 times 0.6) compared to being in the control group. This is in between the high and low MDI estimates for the evaluation of the subsidy pilot activity shown in Table III.1.

Our estimates should be relatively precise for household outcomes for the T&D and subsidy pilot activities. Our estimates will be far less precise for outcomes based on the community, business, and hotel surveys and for the ITT data collected by month in Zanzibar. For the ITT data in Zanzibar this may not be an issue as we expect to see fairly large impacts there. The community, business, and hotel surveys 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 precision in estimates from these surveys should not be critical.

Our ability to detect impacts could be reduced if the activities have little effect on connection rates or if funding constraints cause us to offer subsidies to far less than half of the communities receiving line extensions.

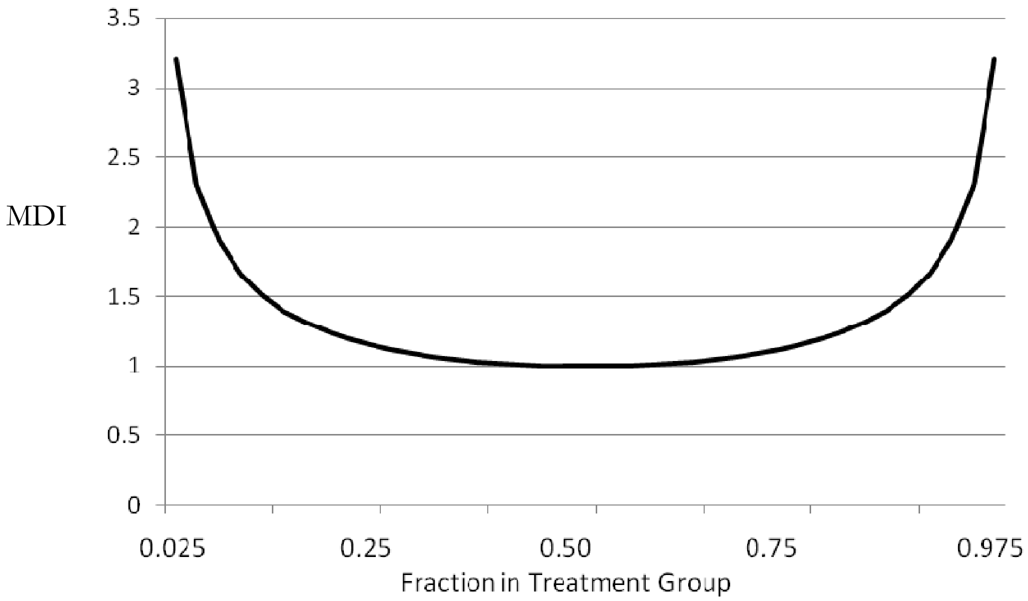
In the T&D example above, we posited that only about 50 percent of all households would connect to the new lines built under the T&D activity. If the fraction of households connecting to the new lines were reduced from 50 to 40 percent, the impacts of the T&D activity will be about 20 percent smaller, holding constant the impacts of actually being connected. To reduce our MDI by 20 percent, we would have to increase the number of communities in our sample by about 56 percent, thereby requiring approximately 100 additional communities in our sample. If the fraction of the households connecting is reduced to only 25 percent, we would have to quadruple the number of communities in our sample. A similar point holds for the subsidy pilot. For example, if the impact of the subsidy on connection rates is reduced from 60 percent to only 30 percent (a shift from a 35 percent connection rate in the control group to a 65 percent rate in the treatment group) then we would expect to see impacts on income only about half as large (i.e. around 2.7 percentile points). In order to cut our MDI from the 4.8 shown in Table III.1 to 2.7 we would need to approximately triple the sample size of communities covered in our subsidy pilot.

In sum, the impacts of the T&D and subsidy pilot activities will depend critically on the actual connection rates of the households. It would be very costly to increase our sample sizes enough to completely offset the risk that these connection rates may be quite low for the T&D intervention communities, or vary little between the treatment and control group for the subsidy pilot.

The fraction of treatment communities MCA-T and MCC will be able to subsidize also matters, but far less than the connection rates among households. In Figure III.1, we illustrate the potential importance of the fraction of communities in the treatment group and demonstrate that the fraction of communities where subsidies will be offered must fall well below 50 percent before the MDI increases substantially from its base (set to 1 in this example). For instance, if the fraction of communities in the treatment group falls to 25 percent (that is, a reduction by half), our MDI would increase by only 15 percent. To summarize, the fraction of households that connect is far more important than the fraction of communities that we can subsidize.

While the fraction of communities we can subsidize does not matter much as long as it remains well above a quarter, we are at substantial risk of going this low because we do not currently know how many households will be eligible for and accept subsidies in each treatment community. Given a budget of $600,000 and a total cost of $280 per subsidy (ready board– and connection cost–inclusive), we estimate that we could provide around 2,142 subsidies, or about 23 per community. Since there may be hundreds of households per community, we could easily end up subsidizing far less than a quarter of communities. Hence, while we do have substantial flexibility with the fraction of communities in our treatment group, we do need to proceed with caution. One way to reduce this risk would be through the systematic omission of larger communities from the subsidy pilot evaluation (though presumably those communities would remain in the T&D evaluation). Once we have completed the baseline survey, we should have a good sense of the number of households in each community. Using the baseline data, we can determine our likely risk of going over budget for the subsidy pilot and the potential benefits of dropping larger communities from the subsidy pilot evaluation.

Figure III.1. Minimum Detectable Impact (MDI) versus Fraction of Sampled Communities in the Subsidy Treatment Group



IV. DATA NEEDS AND SOURCES

Implementing the evaluation designs presented in the preceding chapter will require that we collect and analyze data from diverse sources. In Table IV.1, we summarize the data we will use by identifying the seven required data sets, the evaluation activities in which they will be used, their analytic purpose, and the timing and number of rounds of data collection. In the remainder of the chapter, we briefly discuss the timing of data collection, followed by additional details on each of these data sets in turn.

Table IV.1. Data Sources

Data	Activities	Analytic Purpose	Collection Rounds
National Bureau of Statistics	T&D	To help identify intervention and comparison communities	Once at baseline[a]
Administrative	T&D	To help identify intervention and comparison communities	Once at baseline
Community	T&D	To identify potential comparison communities and to capture community outcomes	Baseline and follow-up
Household	T&D and subsidy pilot	To capture household outcomes	Baseline and follow-up
Enterprise	T&D	To capture enterprise outcomes	Baseline and follow-up
Hotel	Cable	To capture hotel outcomes	Baseline and follow-up
Indicator Tracking Table	All	To capture electricity use, quality, and reliability	Baseline, interim, and follow-up
Qualitative	All	To capture challenges and perceived benefits and costs	Interim

[a] NBS will be collecting another round of Census data in 2012 and possibly also another round of the Household Budget Survey. In theory, those data could help evaluate impacts of the electricity sector project. However, they will likely be collected too soon for that purpose given that the lines will be completed around that time. In addition, they will likely cover relevant outcomes for only a small fraction of the areas in our study—for example, we expect the Census data will only cover 15 percent of the communities--this is the fraction of communities covered by the *long form* in the 2002 census. Similarly, the Household Budget Survey is probably going to be conducted in only a small fraction of communities.

A. Rounds of Data Use and Collection

Data use or collection will occur in three rounds. *Baseline* data will be collected before electrification activities take place. The data will be used primarily as controls in our later analyses and as background to the interim report. *Interim* ITT data will be collected a few months after electrification activities are completed and a few months before the interim report is due so that the interim report can describe early implementation including impacts on electricity use, reliability, and quality. *Follow-up* data will come from surveys conducted at least one year after completion of electrification activities to allow households, businesses, and communities time to take advantage of enhanced access to electricity.[11] This lag should allow time for households, businesses, and community organizations (i.e., schools, hospitals, and water utilities) to pay for wiring and connections and have the wiring completed, for TANESCO to complete the electricity connection, and for these entities to purchase and learn to use electric appliances. The follow-up data will be used for the final report, which will describe overall impacts of the activities.

B. Data Sources

Below, we describe each of the data sources to be used in the evaluations.

1. NBS Data

The NBS data include 2002 census data on households and people, and GPS data identifying the locations of hospitals, schools, roads, and electrical transformers. These data, together with data from the baseline community survey, are being used to identify appropriate comparison communities for the T&D evaluation.

2. Administrative Data

Administrative data obtained from TANESCO and ZECO on the locations of existing and future power lines will also be used to help select intervention and comparison communities for the T&D evaluation.

3. Community Survey Data

The baseline community survey data, which will enable us to improve the match of the intervention and comparison areas for the T&D evaluation, will cover topics related to community well-being and the potential for electricity use. The follow-up community survey will be used to estimate impacts of the T&D activity on such community-level outcomes as prices of goods in the community (especially fuel) and conditions in the local schools, hospitals, and water utilities.

4. Household Survey Data

To estimate the impacts of the T&D and subsidy pilot activities on households, household survey data will be collected during two rounds of in-person interviews. The same set of households surveyed at baseline will be resurveyed at follow-up. We recognize that there will be some attrition from the sample—some households will have moved out of the community, for example. If attrition from the sample is high, we will discuss with MCC the option of sampling a few of these mobile households for follow-up at their new locations. Using data from the baseline survey, we will estimate a multivariate model of attrition from the follow-up survey. If there are significant differences between the intervention and comparison groups for the T&D study or between the treatment and control groups for the subsidy pilot study with respect to predictors in the attrition model, we will weight the follow-up data to eliminate the differences.

[11] The RFP released by MCC in January 2011 suggests less than one year between completion of the Zanzibar cable and the collection of follow-up data. We recommend postponing the collection of follow-up data in Zanzibar to allow hotels more time to take advantage of improved electricity. The interim ITT data can be collected earlier for use in the interim report since we will also be collecting ITT data later for the final report.

To investigate the impact on the intra-household allocation of resources and gender disaggregated impacts, the surveys will ask about the time use of adult household members, specifically, how much time women and men spend cooking and doing other household chores within the household during the day, and working for income outside of the household. Standard of living will be measured using standard batteries of income and consumption questions such as those used by the *Living Standard Measurement Study* to measure wealth. Other general measures of well-being that may be affected through greater and more reliable access to electricity include access to radio and television programming, increased communication with family and friends via cell phones, and perhaps even enhanced food security through improved information about weather-related shocks and changes in market prices.

In addition to the outcome measures noted above, the household surveys will collect information on the quantity of electricity used, reliability and quality of electricity, and the amount spent on electricity and other fuels. The baseline survey will gather information on key background characteristics, including the household's socioeconomic status and the gender and age of each of its members. We will use these background characteristics to create subgroups for the impact analysis and as control variables in regression models to improve the precision of the impact estimates. The baseline survey will also collect contact information so the households can easily be found again at followup; information may include family and business contacts, current place of employment or school, and cell phone numbers.

5. Enterprise Survey Data

Enterprise survey data will be collected either through a panel survey (where the businesses surveyed at baseline are resurveyed in the follow-up) or through repeated cross sections (where a new set of businesses is sampled for the follow-up). We plan to discuss the choice with MCC and MCA-T. One important consideration relating to the decision about which approach to use is the number of businesses likely to close in the time between the surveys. If the rate of business closure is low, the panel survey will provide more statistical power; if the rate of closure is high, conducting a panel may not be possible. An advantage of the repeated cross section is that it will provide more information on the types of large new businesses that appear in these communities. Whatever the decision, we will recognize that any differences between businesses over time or between communities may be a result of changes in existing businesses and/or changes in the composition of businesses. As discussed in Chapter II, Section A, we will use the household survey, aggregated across all communities covered by our survey, to estimate impacts of the energy sector activities on business formation.

Collecting enterprise data will require working with stakeholders to establish definitions of what constitutes an "enterprise," since in many African countries the separation between businesses and households is not clear cut. (For example, shops can be located within households, or households may sustain themselves by farming.) We recommend that for the enterprise survey in the Tanga region, we sample from a list of stand-alone businesses (those that are operated on "premises with fixed location independent from home") in the community. These stand-alone businesses are likely to include medium- to large-size businesses in the study communities, defined in Tanzania as businesses with five or more paid employees. While we expect some overlap in the information collected in the enterprise and household surveys, with a sample of stand-alone businesses in the community the enterprise survey will cover larger businesses, and the household survey will capture small home-based business operations.

As in the household survey, electricity quantity, reliability, and quality of access will be measured in the enterprise survey, as well as the use of traditional energy sources such as kerosene and charcoal. The survey will gather information on the hours that a business is in operation and on its use of electric lighting and other electrical equipment. Survey questions will ask about recent expenditures on capital, including electrical equipment, other equipment or machines, and land or buildings. To estimate worker productivity, the survey will ask about business revenue, the number of employees, and their hours of work. Finally, the enterprise survey will collect data on the characteristics and locations of the businesses to allow the estimation of impacts for different types of businesses and by community characteristics.

6. Hotel Survey Data

Hotel survey data will be collected through a panel survey of 30 hotels and used for the cable evaluation. In the summer of 2010, we collected baseline data monthly for three months. We will also collect follow-up data on a monthly basis for three months. Collecting three months of data (rather than just one) will enable us to better differentiate between typical changes over time and those created by the cable activity. We will use baseline data to describe electricity use before installation of the new cable. Findings from the baseline data will be described in our interim report to MCC and MCA-T. The follow-up data will be collected at least one year after completion of the cable activity to allow us to more completely capture its impacts on the quantity, reliability, and quality of electricity used by the hotels. Results based on those data will be described in the final report.

7. Indicator Tracking Table (ITT) Data

Administrative data from TANESCO and ZECO on the quantity, reliability, and quality of electricity will be used to estimate impacts of the T&D and cable activities on these outcomes. We will conduct these analyses for both the interim and final reports. e. In addition, our survey data at the household, business, and community levels will include parallel constructs to the ITT data collected from TANESCO on energy use, interruption frequency and duration, voltage fluctuations, and substandard voltage level. Thus, we should be able to assess the congruence between the administrative records on power generation and transmission and the actual experiences of households and businesses that are the intended beneficiaries of the activities. Our ability to do this will depend on whether or not the data provided by TANESCO identify the communities affected by the T&D activity and the hotels (as a group) for the cable activity.

8. Qualitative Data

Qualitative data will be collected through semi-structured interviews and focus groups with key community informants, such as business owners, elected officials, school leaders and teachers, health facility managers and staff, TANESCO officials, local government officials, and other community leaders. We propose hiring a local research firm for this data collection task, which will be done once, a few months after completion of the T&D and cable activities and in time for the findings from our analysis of these data to be included in the interim report. During data collection we will ask questions about perceived benefits and uses of electricity and challenges encountered in trying to take full advantage of what is available. Targeted questions will focus on areas of particular concern, such as consumers' limited use of electric stoves, and the reluctance of businesses to invest in electricity even when it is available.

We will pay particular attention to gender differences in perceived benefits and uses of electricity when collecting the qualitative data. We will use this opportunity to explore challenges encountered and how they may differ by gender and be affected and interact with traditional roles that women and men take on in their home and work lives. Electricity may serve as a means to change the relationships between the genders in both positive and negative ways depending on how different household members react to its availability. We will also make use of our qualitative data to enhance our final survey to include outcomes that may not have been considered otherwise.

V. ANSWERING KEY EVALUATION QUESTIONS

The energy sector evaluation will be a success only if it thoroughly addresses the key research questions that guide the study. We explain below how the results of our evaluations of the energy sector project activities will be used to address the seven main research questions.

Research question 1—impacts on implementation measures: What is the impact of the project on access to electricity, energy quality, and energy reliability?

To address this question we will use ITT data on electricity access, energy quality, and energy reliability obtained from TANESCO and ZECO as well as data from the surveys. We do not expect to have the ITT data for individual households and businesses but do hope to be able to obtain such data by geographic area (for the T&D activity) and by month for the cable activity from TANESCO and ZECO. We do expect to have individual data on households and businesses from the survey data. We plan to use DID to analyze these data for the T&D activity, random assignment for the subsidy pilot activity, and pre-post for the cable activity. We will use our results to estimate the impacts of each activity on the numbers of households and businesses connected and the quality and reliability for their service.

Research question 2—impacts on outcomes: Does access to electricity improve intermediate and long-run outcomes for households, communities, and businesses? If impacts are detected, what are the magnitudes of these impacts?

To address this question we will use survey data collected from households, communities, and businesses covering the appropriate outcomes. We will use the same methods used to answer research question one, but the focus will be on the intermediate and long-term outcomes identified in our conceptual framework, rather than on the short-term outcomes addressed in research question 1.

Research question 3—benefit-cost: Is the project warranted based on the benefit-cost analyses?

Electrification projects such as those being undertaken in Tanzania use resources that could have been used for other improvements. Because prior research has led to mixed conclusions about the cost-effectiveness of these programs (IEG 2008), an important element of the evaluation is to assess whether the interventions yield the economic improvements that were anticipated. Findings generated from this evaluation, together with estimates of the costs of the interventions from MCA-T, will be used to calculate the realized ERR of the program. The realized rates can then be compared against the projected ERRs to test whether the original projections differ from the realized outcomes, and to investigate which components of the original calculations contribute to any observed differences.

Research question 4—analyzing benefit of rigor: Would a simple before-and-after comparison of beneficiaries yield the same results as DID or random assignment?

To examine whether a before-and-after comparison would have yielded results similar to our estimates based on DID or random assignment methodologies, we will estimate impacts using pre-post methods for the T&D and subsidy pilot activities and then compare those results to the estimates obtained using our rigorous methods. We will adjust our statistical tests of the differences to account for the correlations between the rigorous and non-rigorous estimates that result from their being estimated using the same data. This adjustment will improve our ability to estimate the differences if the estimates are positively correlated (which is what we expect) compared to assuming that they are drawn from independent samples.

Research question 5—analyses by gender and other subgroups: Do the impacts vary by gender, age, and income?

When sample sizes allow, we will estimate impacts separately for subgroups based on gender, age, and income. Since gender is a strategic priority at MCC and MCA-T, we will be specifically focused on methods designed to disaggregate results by gender and make optimal use of gender sensitive evaluation and data collection strategies relevant to the context of the energy sector project and Tanzania. We will disaggregate results by gender in at least two different ways. First, impacts will be estimated on outcomes captured in the household survey that separate out female and male household members or employees. For example, the household survey will ask questions regarding earned income and time allocation separately for male and female adults in the home and the enterprise survey will ask about male and female owners of businesses and employees, as we expect that impacts for these outcomes may differ significantly by gender. We expect that the sample sizes for the household analyses should be similar to those used for outcomes that are not broken down by gender, as we expect that most households will have male and female adults. Second, we recommend estimating impacts separately by the gender of the head of the household or enterprise. However, if there are few female-headed households or large enterprises, the statistical power for those analyses may be poor, and those analyses may therefore remain exploratory in nature.

We also plan to use gender sensitive evaluation and data collection strategies. For example, we will ask a few questions to both males and females (for example, on joint income); so we will be able to look at some results separately based on the gender of the respondent. This is important given evidence that males and females answer some questions differently (e.g., Bardasi et al. 2010). We will also make sure that the interviewers are trained to remain sensitive to appropriate cultural norms when interviewing male and female respondents and, when possible, to interview them separately to avoid the possibility that one respondent is influencing the responses of the other. Finally, we will use the qualitative data collection component of our evaluation to explore how gender roles may affect electricity use and be affected by access to electricity in unanticipated ways that may both enable us to enrich our final survey and better understand the results we see in our quantitative analyses.

We will also be glad to explore with MCC and MCA-T more in-depth ways of examining the gender aspect of the impacts of electrification. Such examination could range from additional quantitative assessments, to focus group discussions of particular issues that may be relevant for gender and electricity, such as the education of children, to a policy brief on gender issues as one of the outreach materials we create.

Research question 6—unintended consequences: Are there unintended results of the program?

We will include in our surveys several variables designed to capture unintended consequences of the program, both positive and negative. On the positive side, we will measure the building of other types of infrastructure (such as roads and water) to determine whether these types of projects are more prevalent when electricity projects are implemented. On the negative side, we will look at accidents, such as fires and injuries caused by electricity, as well as potential changes in time utilization, such as time spent watching TV instead of studying (to see if access to electricity results in some students doing less homework). We will also include open-ended questions in qualitative component of our evaluation to help capture other potential unintended consequences, both positive and negative.

Research question 7—lessons learned: What are the implications of the evaluation findings?

We will describe the implications of our results for the roll out and implementation of future impact evaluations and for long-term policy decisions. For example, challenges we faced doing the regression discontinuity and identifying households within 30 meters of where the lines will be built could influence future evaluations of electricity sector projects, as discussed in Appendix A. Similarly, findings from the subsidy pilot might suggest the need for additional work to determine the optimal subsidy level, and the estimated impacts of the subsidy pilot on connection rates may have relevance for ongoing discussions of changes in the connection fees charged to customers. Finally, ERRs based on the T&D line extension study may help influence the government of Tanzania's decisions about how much to invest in electricity as compared to other infrastructure, such as roads and water.

VI. TIMELINE AND WORK PLAN

The success of the energy sector evaluation depends on having a schedule and work plan that ensure timely completion of various evaluation tasks, responsiveness to MCC and MCA-T's priorities, and attentiveness to cost control. In this chapter, we discuss the timeline and a work plan for evaluation of the three energy sector activities—T& D activity, subsidy pilot activity, and cable activity—that aim to achieve these goals. A combined schedule for the evaluation tasks across these activities is presented in Appendix B.

Table VI.1 shows our timeline for the evaluation of the T&D activity. The installation of new electricity lines under the T&D activity will occur in phases over a period of about a year. Ideally, our baseline survey would be conducted before any households are connected to the new lines and the follow-up would occur long after so full impacts of the interventions could be realized. Because of the long time period during which the T&D activity is being implemented, it appears that we may not be able to achieve either of those goals easily. The current schedule for the line extension activity suggests that some households will be connected by August 2011, about two months before our baseline survey ends. Fortunately, however, most households will not be able to connect by that time. Indeed, we expect that only about 13 percent of the line extensions will be complete by August 2011 and it will probably take households a few more months after that to connect. Hence, we expect very few households will have connections during our baseline survey. In addition, the survey firm may be able to time its data collection activities to help ensure that the surveys for almost all households are conducted before connections are made. Therefore, while our baseline will not be perfect, we anticipate it will be very close to the ideal.

Table VI.1. Timeline for Evaluation of T&D Activity

Date	Event
March 2011	Start of baseline survey
August 2011	Energization start
November 2011	End of baseline survey
September 2012	Connections completed
January 2013	Collect interim ITT data
January 2013	Collect qualitative data
March 2013	Interim report
September 2013	End of Tanzania compact
September 2014	Start of follow-up surveys
December 2014	Collect Final ITT data
March 2015	End of follow-up surveys
June 2015	Final report

At the other end of the schedule, we would like to be able to produce a final report about six months before the end of the five-year compact that Tanzania has with MCC and MCA-T. Unfortunately, the current schedule suggests that connections will still be made as late as September 2012. Since we want to give households at least one year to connect and learn to take advantage of those connections, this means that the follow-up survey could not start until September of 2013, when the compact ends. Since the follow-up survey will take a few months to be completed and additional months will be needed to analyze the data and write a report, the final report based on the results from the follow-up data analysis will not be available until long after the compact is finished. This is one reason we are planning to produce an interim report using preliminary data available, such as ITT data on electricity use from TANESCO and qualitative data collected from focus groups and interviews with key community informants on perceived benefits and challenges associated with using electricity.

Table VI.2 shows the timeline for the subsidy pilot activity evaluation. The timing of the subsidy pilot is somewhat more complicated than the T&D evaluation because we have a budget constraint that limits the number of subsidies we can provide, but we do not know how many households will accept the subsidy offer. Therefore, it is difficult to determine the correct number of communities (subvillages or urban neighborhoods) to put in our treatment group. If we include too many communities, we risk going over budget for the subsidies. If we include too few we risk having a study that is under-powered. To help minimize both of these risks we are recommending that the subsidy pilot activity be implemented in two stages. In the first stage, a small randomly chosen subset of communities will be offered the subsidies. Then, based on the observed connection rates in those communities, we will decide how many communities we can offer subsidies to in a second stage. The result is that we hope to be able to make optimal use of the subsidy budget available by getting as many treatment communities as possible into our study while minimizing the risk of going

Table VI.2. Timeline for Evaluation of Subsidy Pilot Activity

Date	Event
November 2011	End of baseline survey
November 2011	Phase I random assignment
November 2011	Subsidies offered
December 2011	Households accept subsidy offer
January 2012	Subsidized connection eligibility confirmed
January – February 2012	Mathematica analyzes phase I data on connection rates
February – March 2012	Phase II random assignment
March – April 2012	Subsidies offered
April 2012	Households accept subsidy offer
May 2012	Subsidized connection eligibility confirmed
September 2012	Line extensions completed
January 2013	Collect interim ITT data on electricity connection rates
March 2013	Interim report covering connection rates
September 2013	End of Tanzania compact
September 2014	Start of follow-up surveys
December 2014	Collect Final ITT data
March 2015	End of follow-up surveys
June 2015	Final report

over budget. As shown in Table VI.2, this means the subsidies will be delayed for a fairly long period of time for many households. However, since the final connections are not expected to be made until September of 2012, we do not expect that the subsidy pilot will delay the follow-up survey.

Table VI.3 shows the timeline for the cable activity evaluation. This is quite similar to that for the evaluation of the T&D activity. We will produce interim estimated impacts of the cable in time for the interim report that will be produced in March 2013, six months before the end of the compact. In theory we could try to collect the follow-up hotel survey data in time for the interim report also. However, we do not recommend this, as it would likely mean either collecting this hotel survey data in the winter months (November-January) in which case it would not match as well with the baseline data collected during the summer, or having insufficient time to carefully analyze and write about the results. Consequently we are proposing to include results for the hotel follow-up survey in our final report in 2015.

Table VI.3. Timeline for Evaluation of Cable Activity

Date	Event
Summer 2010	Baseline survey
Summer 2012	Cable completed
January 2013	Collect ITT and qualitative data
March 2013	Interim report
September 2013	End of Tanzania compact
Summer 2014	Follow-up hotel surveys
December 2014	Collect final ITT data
June 2015	Final report

VII. CONCLUSION

In this report, we have presented our plan for conducting a rigorous evaluation of the energy sector project being funded by MCC in Tanzania. Baseline data collection for this work is underway, and most key decisions regarding the evaluation design have been made. However, there are a number of potential challenges that may arise in the course of carrying out the evaluation. In Section A of this final chapter, we identify these issues and propose approaches to resolve them. Additional, unforeseen challenges may arise in the future. If so, we will deal with them using approaches similar to those described below, maintaining as much rigor as possible while continuing to avoid putting undue burden on the entities implementing the sector activities. In Section B of this chapter, we discuss the next steps for the evaluation.

A. Potential Challenges and Proposed Responses

1. T&D Evaluation Issues

There are a number of potential issues with the evaluation of the T&D activity. These relate to activity-induced household mobility, within-community confounding, between-community confounding, cross-over, late changes in the subproject list, identifying subvillages within each village, and the timing of the follow-up survey.

Mobility of households could bias DID impact estimates if the introduction of electricity to a community were to affect people's decisions to move into or out of that community and if it is difficult to follow households that move. Site visits in 2009 suggested that there may be substantial household mobility when electricity is introduced. Some people may move into a community with new electrical lines to take advantage of the electrical connections, and others may leave, perhaps to take advantage of increased land prices. We recommend excluding the in-migrants from the study because we have no way to identify appropriate comparison households for in-migrants and also because we have no baseline data for these households. In-migrants in the comparison communities are likely to differ since the intervention communities have a key asset (electricity access) that is lacking in the comparison communities. We also recommend caution regarding the inclusion of out-migrants in the study because doing so could greatly increase data collection costs. As an alternative to the full inclusion of out-migrants in the study, we will determine how many baseline respondents move out of their original communities and ask non-movers about their neighbors who migrated out and their decision to leave in the follow-up survey. Dinkelman (2008) did find evidence that electrification induces mobility, though in her study it was not sufficient to have a major impact on her estimates of the benefits of electrification. We will assess the mobility situation at followup. If we find high mobility, along with evidence that it was induced by electrification, then we will explore the methodological approaches that are available in the evaluation literature (e.g., McConnell et al. 2008; Rubin 2006; Lee 2005) to address the issue, including surveying a randomly chosen subset of the out-migrants.

Within-community confounding could occur if, in some of the study locations, multiple government, MCC, or other donor-funded water and road interventions are implemented during the same time that the MCC-funded electricity projects are being implemented. In this scenario, it might be problematic to estimate the direct impacts of increased access to electricity; however, DID could still be used to estimate the overall impact of the combined interventions. We expect that there will also be interest in estimates of the direct impacts of the electricity projects without the indirect impacts of roads and water. It would probably not be possible estimate the direct impacts of the electricity projects rigorously but we would explore alternative methods. One approach would be to exclude communities with road and water projects from the impact analysis. To maintain the balance between the intervention and comparison groups, the communities that had been matched to those with road or water projects would also be excluded from the analysis. Another approach would be to include identifiers for road and water projects as control variables in the regression models that we will use to estimate impacts. With either of these two approaches, we would be concerned that the resulting estimates would be less rigorous than the main estimates without such adjustments, but of course the main estimates would suffer from the confounding problem. Even if we do not use the alternative methods, we will collect data on concurrent road and water interventions during the baseline and follow-up surveys and use that information to help with the interpretation of our main impact estimates.

Between-community confounding could arise if some nonresidents of the intervention communities receive education, health care, water, and other services from those communities. This could result in underestimation of the full benefits of the electrification of schools, health care facilities, water systems, and other community facilities. While the DID methodology will enable us to estimate the benefits of the electrification of shared facilities within the same community, it will probably not incorporate the full benefits of these shared facilities since at least some individuals residing outside of the communities receiving the new MCC line extensions will benefit from those facilities. To assess the potential magnitude of this issue, we recommend the collection of data on the prevalence of the use of such facilities by nonresidents of the study communities during the follow-up survey.

Crossover is another common analytic challenge in random assignment evaluations that could arise in the T&D evaluation. Since we are selecting intervention and comparison communities before the electric lines are built, it is conceivable that some of the intervention communities may end up not receiving new electricity lines, while some of the comparison communities may receive new lines. If a substantial fraction of communities crossover between the intervention and comparison groups in this way, then we could use the instrumental variables (IV) method to estimate the impacts of being in a community that actually receives a new electricity line, with the original intervention status as the instrumental variable.[12] The IV method is preferable to simply dropping from the analysis the communities that crossed over. The problem with the latter approach is that remaining communities in the intervention and comparison groups might no longer

[12] The instrument variables method can be thought of as a two-stage regression model. In the first stage we regress the final intervention status (1 if the community actually received electricity from MCC funded lines, and 0 otherwise) on the original intervention status (1 if the community was supposed to receive a new line, and 0 otherwise) and the baseline control variables we are planning to use. We then use the results from this first stage regression to create a predicted probability of being in a community that actually received new line extensions. In the second stage we regress the household, community, or business outcomes of interest on the predicted probability as well as the baseline controls. The coefficient estimate on the predicted probability is the estimate of the impact of being in a community that actually receives a new line extension.

We recommend this method because our intervention and comparison areas were chosen to be similar. Thus, the original intervention status should not affect later outcomes except through its impact on being connected. This suggests that the original intervention status would be a valid instrument. In contrast, there is good reason to believe that the cross-over decision is endogenous, that is, the communities that cross-over are not the same as those that do not. This means that the final electrification status of a community, which is determined in part by the cross-over, could be correlated with baseline differences that were not controlled for in our design. Instead of using the instrumental variables approach, one might want to estimate impacts of electricity by comparing outcomes for communities with electricity to those without at the time of the follow-up survey. Since these communities include those that crossed over, the resulting differences could be affected by baseline differences and would not, therefore, provide valid estimates of the impacts of electricity for those communities.

be equivalent, thus the advantage of having a matched comparison group would be compromised and the impact estimates could be biased.

Late changes could occur to the list of areas receiving MCC-funded line extensions. In particular, it is possible that some of the subprojects identified for line extensions may be dropped due to unforeseen developments. In order to accommodate this risk, we conducted the first phase of propensity score matching for all 337 communities that are currently slated to receive line extensions. We then randomly sorted these communities and selected the first 182 to participate in the evaluation as members of the draft intervention group. If any of these intervention communities must be dropped from the evaluation due to cancellation of their line extensions, we could draw replacements from the original group of 337 communities. However, this replacement procedure would be feasible only for intervention communities that are dropped from the evaluation prior to the completion of the baseline community survey.

Identifying a subvillage within each intervention village may be non-trivial. In each village, we will develop a household listing and conduct a household survey in a selected subvillage, instead of the entire village.[13] We had hoped to obtain information from TANESCO on the subvillage within each intervention village that has the most households likely to have access to the new lines. However, after discussions with MCA-T energy sector team, we concluded that TANESCO is unlikely to be able to provide this information consistently for all intervention villages. Consequently, we plan to ask NRECA to collect similar information from leaders of the intervention villages during the community survey. Specifically, the NRCEA data enumerators will ask the village leaders about whether new lines will be constructed in a subvillage and the fraction of households in each subvillage that are likely to have access to electricity through the new lines. The data enumerators will then select the subvillage within each village with the largest fraction of households that will have access to electricity from the newly extended lines. If multiple subvillages are reported to have the same fraction of households having access to the new lines, the data enumerators will randomly select one of them. In the process of collecting the required subvillage information from the village leaders, the data enumerators may refer to the sketch maps and line diagrams they will have with them, and may need to probe to obtain the appropriate information from the village leaders.

Optimal timing of the follow-up survey may be challenging to achieve. On the one hand, it might make sense to delay the survey in order to give households and businesses in the intervention communities more time to take advantage of the new power lines. Indeed, recent evidence suggests that it might take about two months for businesses to get connected in Tanzania (World Bank, 2011). On the other hand, the longer we wait, the more likely it is that some of the comparison communities will get access to the lines. If that were to happen, expected impacts would be reduced. Given these considerations, it might make sense to closely monitor connection rates with TANESCO and to try time the survey so that the difference in connection rates between the intervention and comparison communities is at its highest possible level. This would optimize our ability to detect the impacts of the intervention. However, the follow-up survey fielding period is projected to be six months, so if we were to wait until the gap in connection rates between intervention and comparison communities is greatest, we might find that the gap will have narrowed in the latter part of the fielding period. Therefore, we will also have to take into account the fielding period in deciding when to implement the follow-up survey.

[13] We do not have to select sub-units in urban neighborhoods/*mtaa* because a *mtaa* is the smallest administrative unit in urban areas. Thus, household listing and household survey will be carried out in an entire *mtaa*.

2. Subsidy Pilot Evaluation Issues

Most of the issues that may arise for the evaluation of the T&D activity are also relevant for the evaluation of the subsidy pilot activity. In addition, the subsidy pilot evaluation has two unique issues: (1) The number of communities in which subsidies can be provided is uncertain; (2) the most cost effective method for installing subsidized connections is not currently known.

The number of communities in which subsidies can be provided is unclear for a number of reasons. While MCA-T has budgeted approximately $600,000 for the subsidy pilot, it is not yet known whether these funds must cover the cost of implementing the subsidy pilot activity, in addition to the cost of the subsidies themselves. Moreover, the cost of each subsidized connection is also not known. Finally, we do not know how many households will accept the subsidy offer in each treatment community. Greater clarity on the number of communities where the subsidies can be offered will be needed at least one month before the start of Phase I random assignment for the subsidy pilot, in order to determine the optimal number of communities to include in that phase. If the total number of communities in which we can afford to subsidize electrical connections is smaller than is currently expected, then we may reduce the number of treatment communities in Phase I to ensure that we do not go over budget in that phase of the study. We assume that MCC will provide whatever information it can on available funding at that time (one month before the start of the Phase I random assignment). The evaluator can then use this information to determine the optimal size of the Phase I group. A larger number of communities in Phase I will give a more precise estimate of the likely subsidy take up rate in Phase II. A smaller group will reduce the risk of going over budget in Phase I.

Installing the subsidized connections most cost effectively will be challenging. One way to minimize the installation cost would be to install the connections at the same time that the lines are built, so that one firm can do both tasks at the same time. It may be possible to do this in selected communities where connections can be installed quickly. This could save resources that could then be used to extend the subsidy to a larger number of households. However, we do not expect that it will be possible to install connections concurrently with extending the power lines in many communities without reducing the quality of the evaluation. In order to install connections simultaneously with line extensions, we would have to postpone the household survey, and thereby lose our ability to control for differences in household baseline characteristics using measures from the household survey. This would lead to less precise impact estimates from follow-up survey data. This would also compromise the credibility of the DID method for evaluating the T&D activity because it would mean that we would have a much weaker baseline measure to use to test how well our intervention and comparison communities were matched.

B. Next Steps

The most immediate next steps in the evaluations of the T&D activity and the subsidy pilot activity are the following:

- Deciding what questions about subvillages to include in the community survey so that the resultant data will support our identification within each village of a subvillage that will have substantial access to the new power lines.
- Finalizing the variables that must be captured by the household and enterprise baseline surveys, providing them to NRECA, and then working with NRECA, MCA-T, and MCC to develop the survey instruments.

- Determining how many electrical connections can be subsidized with the existing funds available from MCC and MCA-T for the evaluation of the subsidy pilot activity. This will depend on the total funding available, the cost of other related activities (such as implementation of the subsidy pilot), and the cost of a typical connection.
- Deciding whether the enterprise survey should be a two-wave panel survey or two separate cross-sectional surveys. If it is going to be a panel survey, then we need to determine whether we will try to follow businesses that move. Conducting a panel survey would provide more precise information on changes over time for the non-movers and to measure the rate at which businesses move out of an area. Conducting two separate cross-sections would enable us to measure new business formation. These decisions do not have to be made until long after the baseline survey has been completed. Indeed, we might consider conducting a mini-survey in a few intervention villages shortly before the follow-up enterprise survey, to determine what fraction of businesses move and what fraction are new since the baseline survey.[14] If those fractions are small then we might stick with the panel survey. If they are large then we might recommend using repeated cross-sections or a mixed method (partly cross-sectional and partly panel). Regardless, if a panel survey is being contemplated, then we will need to collect enough contact information in the baseline survey to follow businesses, should they move.

[14] This could also be done during the interim data collection phase when we would collect qualitative data, but that might be too soon to provide relevant information for the follow-up survey.

REFERENCES

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

Bardasi, Elena, Kathleen Beegle, Andrew Dillon, Pieter Serneels. "Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania" *Policy Research Working Paper* 5192, Washington, DC: The World Bank, January 2010.

Chaplin, Duncan, Minki Chatterji, and Denzel Hankinson. "Final Design Report for the Millennium Challenge Corporation (MCC) Impact Evaluation Design & Implementation Services—Tanzania Energy Sector Project." Washington, DC: Mathematica Policy Research, July 2, 2009.

Dinkelman, Taryn. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." Princeton, NY: Princeton University, November 2008.

Ellis, Amanda, Mark Blackden, Jozefina Cutura, Fiona MacCulloch, and Holger Seebens. *Gender and Economic Growth in Tanzania: Creating Opportunities for Women*. Washington, DC: The World Bank, 2007.

Hankinson, Denzel, Lauren Pierce, Duncan Chaplin,and Arif Mamun, "Subsidy Pilot Memo." Washigngton, D.C.:Mathematica Policy Research, February 4, 2011.

Independent Evaluation Group. "The Welfare Impact of Rural Electrification: A Reassessment of the Costs and Benefits." *An IEG Impact Evaluation*, The World Bank, Washington, DC, 2008.

Lee, D.S. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." Cambridge, MA: National Bureau of Economic Research, 2005.

McConnell, Sheena, Elizabeth Stuart, and Barbara Devaney. "The Truncation-by-Death Problem: What to Do in an Experimental Evaluation When the Outcome Is Not Always Defined." *Evaluation Review*, vol. 32, no. 2, April 2008.

Millennimum Challenge Corporation (MCC). "Millennimum Challenge Compact between the United States of America, acting through the Millennium Challenge Corporation, and the Government of the United Republic of Tanznia, acting through the Ministry of Finance," Washington, D.C, February 2008.

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.

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. B. "Causal Inference Through Potential Outcomes and Principal Stratification: Application to Studies with "Censoring" Due to Death." *Statistical Science,* vol. 21, no. 3, 2006, pp. 299–309.

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

Sweco. "Makambako Substation and Network Extension in Njombe, Tanzania". Draft Report Submitted to TANESCO. October 2008.

The Revolutionary Government of Zanzibar. Zanzibar Strategy for Growth and Reduction of Poverty. Zanzibar, Tanzania, January 2007.

World Bank. *Doing Business 2011 - Tanzania: Making a Difference for Entrepreneurs.* Washington, DC: The World Bank and the International Finance Corporation, 2011.

World Bank. *World Development Indicators 2009.* Washington, DC: The World Bank, 2009.

APPENDIX A:

ALTERNATIVE EVALUATION APPROACHES CONSIDERED FOR THE LINE-EXTENSION STUDY

We considered many variations of the evaluation design plans described above. In this appendix we briefly discuss two major variations that were considered for the T&D evaluation—using a "Regression Discontinuity" method to analyze the data, and using a "Walk the Line" method to identify households that were likely to get access to line extensions in both the intervention and comparison communities. Following is a brief discussion of both of these issues along with our rationale for not using them.

A. Regression Discontinuity

Our original plan for the line extension study was to use the Regression Discontinuity or "ordered list" method to estimate impacts of extending lines to new communities in the six D&E regions (see Chaplin et al, 2009). To employ this method, we would need a list of potential line extension subprojects. The determination of which subprojects are implemented must be based almost entirely on one variable, such as the revenue/capital expenditure ratio. To illustrate the method this list could be sorted in the order of that variable. In order for the method to work, almost all potential subprojects with values above some cut-point receive MCC-funded line extensions and almost all of the others do not. During a trip to Tanzania in February 2009, we discovered two major problems with using RD for the line extension study. The first issue is that funding decisions were not solely based on the revenue/capital expenditure ratios. Rather, other factors played a major role. The second is that the sample size of unfunded subprojects below the cut-point appears to be very small. In light of this, the evaluation methodology was changed to DID, which also required changes in the data collection activities and resulted in delays in the implementation of the baseline survey.

B. Walk the Line

In developing this design, we considered limiting our sample to households that would be close to the planned lines. This is challenging for two reasons. First, we want to conduct the household survey before the lines are built in the intervention communities. This means that we cannot easily identify which households will be close to the lines there. Second, we do not have a good way of identifying which households in the comparison communities would have been close to an existing line, had such lines been built there.

For the comparison communities we considered asking TANESCO staff to identify where lines would go in those communities, if lines were to be built. During a meeting in Tanzania in November 2010 (when the survey firm, NRECA, was present) it was decided that this would likely be a very costly exercise. Consequently, we decided to change our plan and to instead focus on the subvillage with the greatest number of expected customers, rather than try to identify which households would be close to a new line, if one were to be built.

APPENDIX B:

SCHEDULE OF ENERGY-SECTOR EVALUATION TASKS

Figure B.1. Schedule of Energy Sector Evaluation Tasks by Year and Quarter

Activity/Task	2011			2012				2013				2014				2015		
	Apr-Jun	Jul-Sep	Oct-Dec	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	Jan-Mar	Apr -Jun	Jul-Sep
Planning	D	D																
T&D Activity						A												
Subsidy Pilot Activity						A												
Community Surveys	B													F	F			
Household Surveys		B	B											F	F	F		
Enterprise surveys		B	B											F	F	F		
Cable Activity						A												
Hotel Surveys													F	F				
Qualitative Data Collection								I										
ITT Data Collection								I							F			
Interim Report								D										
Final Report																	D	
Compact End										End								

A=Activity Completed, B = Baseline Data, D = Major Deliverable, F = Follow-up Data, I = Interim Data

T&D Activity indicates that line extensions have been completed.

Subsidy Pilot Activity indicates that connections have been made.

Cable Activity indicates that the new cable has been connected.