Evaluation Design
MCC Connectivity Project in El Salvador

The Northern Transnational Highway

By: Máximo Torero[1]

International Food Policy Research Institute
Consortium led by Social Impact

July 2013
Revised: September 2013, February 2014

[1] Máximo Torero is the Director of the Markets, Trade and Institutions Division of the International Food Policy Research Institute (IFPRI), m.torero@cgiar.org . This report was done with the collaboration of Eduardo Nakasone and Miguel Almánzar.

I. List of Acronyms

CGE .... Computable General Equilibrium

DIGESTYC........ Dirección General de Estadísticas y Censos - El Salvador National Statistics

EHPM........ Encuesta de Hogares de Propósitos Multiples – Multi-purpose Household Survey

ERR .... Economic Rate of Return

FOMILENIO........ Fondo del Milenio (Millenium Fund)

GDP .... Gross Domestic Product

GIS .... Geographic Information System

GOES........ Government of El Salvador

HDM .... Highway Development and Management Model

IHSN........ International Household Survey Network

MCC........ Millenium Challenge Corporation

MDE........ Minimum Detectable Effect

NCR........ Network of Connecting Roads

NTH .... Northern Transnational Highway

RD........ Regression Discontinuity

II. Table of Contents

III. List of Tables

IV. List of Figures

1. Introduction

In November of 2006, the Millennium Challenge Corporation (MCC) signed a five-year, $461 million Compact with the Government of El Salvador (GOES) to improve the lives of Salvadorans through strategic investments in education, public services, agricultural production, rural business development, and transportation infrastructure. The Government of El Salvador set up a management unit called FOMILENIO to implement the Compact from September 2007 to September 2012.

Social Impact was contracted by MCC to conduct an impact evaluation of the Compact's connectivity project. While this project initially consisted of a network of connecting roads (NCR) and the Northern Transnational Highway (NTH), because of a significant increase in construction costs and the existence of other interventions designed to connect roads, the NCR was not built. Thus, this evaluation only focuses on the Northern Transnational Highway.

The impact assessment will combine two parallel approaches. The first approach takes advantage of the sequence in which the different segments of the NTH were constructed by combining a regression discontinuity and a pipeline design. The discontinuity is created by the boundaries of the different segments while the pipeline design is created using the different construction dates. The second approach exploits variations in the intensity of treatment. These variations result from the fact that over time, the NTH will provide different degrees of accessibility to the households located along the road. As a result, a continuous treatment approach will be implemented. Both methodologies exploit the panel structure of the data to measure the change in household incomes within the "area of influence," defined as the area within 30 minutes of the NTH via existing means of communication. Additional outcomes that will be evaluated include: the reduction of transportation costs and transportation time, land values, access to public services and their impacts on health and education outcomes, changes in labor allocation between farm and non-farm activities, and differentiated gender effects of road improvements.

This document updates the initial evaluation design report to account for the changes that occurred during the implementation of the NTH, to gauge how these changes affected the original design, and to determine what can be done to improve the design in response to these changes.

2. Overview of the Compact and the Intervention Evaluated

The El Salvador Compact began in September 2007 and ended in September 2012. The Compact consisted of three projects; these projects had the collective goals of stimulating economic growth and reducing poverty: productive development ($68 million), human development ($89 million), and connectivity ($269 million). The human development project consisted of two activities: education and training activity and community development activity. The community development activity consisted of three sub-activities: rural electrification sub-activity, community infrastructure sub-activity, and water and sanitation sub-activity.

The connectivity project consisted of one main activity – the rehabilitation and improvement of the Northern Transnational Highway (NTH), a vital transport artery that runs throughout the region as well as into neighboring Honduras and Guatemala. The goal of the NTH is to increase the Northern Zone's access to and connection with markets and the larger regional and national economy.

2.1. Program Logic

The overall logic of the Compact was to improve the lives of Salvadorans in the Northern Zone. As such, the Compact combined infrastructure development with technical assistance aimed at connecting northern El Salvador with the rest of the country, helping to create opportunities for the region's residents through increased access to markets through the east-west highway, increased access to electricity through expansions of the electrical grid and distribution of solar panels, increased access to water and sanitation facilities to decrease disease in the region, and other interventions in education, agriculture, and other productive activities.

2.1.1. Compact-level

The Northern Zone of El Salvador contains half of El Salvador's poorest municipalities and suffered more damage from the country's internal conflict during the 1980s than any other region. Economic and social indicators in the Northern Zone are currently worse than the national average: In 2011, 48.4 percent of households in the Northern Zone were poor, compared with the 40.6 percent national estimate; 18.7 percent of households in the region lived in extreme poverty in 2011 compared with 11.2 percent at the national level. Human capital development is also lower in this region than in any other. The average level of schooling in El Salvador was 6.2 years in 2011, while the average in the Northern Zone was only 4.7. The percentage of illiterate people in the Northern Zone was 21.9 percent in 2011 versus a 12.8 national average[2]. The goal of the Compact was to reduce rural poverty by increasing regional economic growth through a five-year program of strategic investments and technical assistance in various sectors.

2.1.2. Project-level

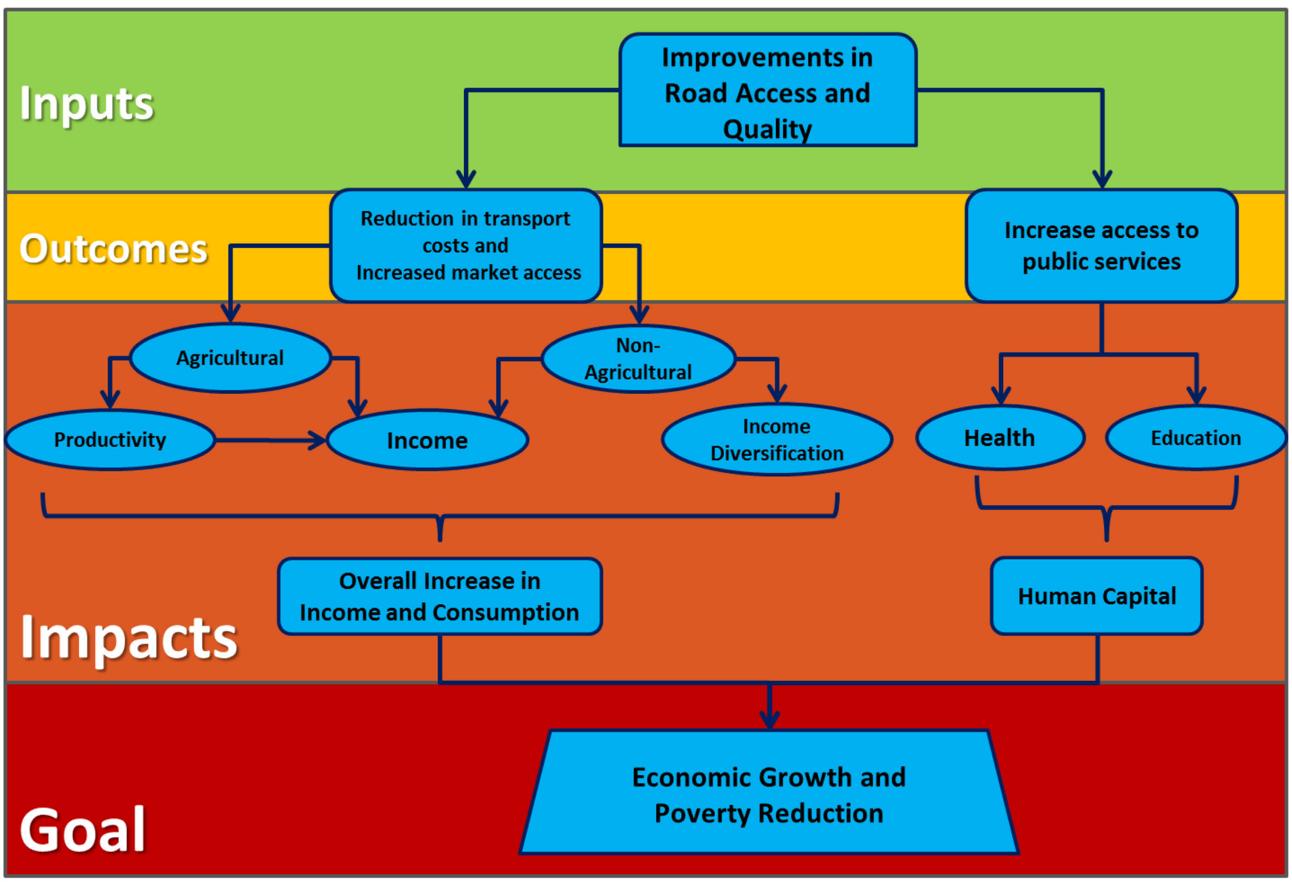
The NTH serves as a transportation artery within the Northern Zone and could potentially improve international connectivity with Honduras to the east and Guatemala to the west. The project constructed, improved, or rehabilitated 280.7 km of the NTH, allowing the highway to provide contiguous and reliable access to communities in the Northern Zone, as well as to main transportation corridors. This should enable the Northern Zone to participate more fully in the national and regional economy.

Reduction of the transportation costs within the Northern Zone to the rest of the country and to neighboring countries could facilitate the access to markets, promote territorial development, increase productive use of land, and attract new investments. The increase in accessibility could also improve access to health and education services. Together these effects are expected to cause an overall improvement of welfare of beneficiary households.

[2] Source: DIGESTYC from national household survey 2011 and 2012 Dirección General de Estadística y Censos (2011, 2012)

The impact pathways below describe the expected causal chain of events leading from project activities to outputs, to changes in the target population, and to the achievement of project objectives. Figure 1 shows the impact pathways of the NTH construction. The figure shows how we expect the road improvements to affect the livelihood of the poor in the area. First, through the income-market access pathway, increases in access through lower times and lower cost of moving products to existing markets, are expected to promote agricultural productivity and participation in non-agricultural activities, by the availability of better and cheaper inputs for agricultural activities and the increased demand for non-agricultural labor from new and more accessible existing markets. These changes imply income flows that are more diverse and perhaps less volatile promoting resilience in those that exit out of poverty. Second, improvements in the roads could also improve the access to existing health services, education services and other existing public infrastructure. This can increase the use of health facilities and school enrollment, increasing the human capital of a wide range of the population, specifically vulnerable sector (children, women and elderly). These two pathways compound their effect to achieve the objective of promoting inclusive and sustainable growth in the region.

FIGURE 1 NTH IMPACT PATHWAYS



2.2.ERR and Beneficiary Analysis

The Economic Rate of Return (ERR) measures the effectiveness of a program by contrasting the discounted flows of costs and benefits of a specific intervention. The costs are comprised of any initial investment and any required maintenance expenditures throughout the course of the program. The benefits are determined by the gains of the population affected by the project. The updated mid-term (2010)review ERR of the connectivity project was 16 percent over 20 years, including Compact administration costs; this was revised to 23.9 percent and compact closing in 2012 (MCC, 2012).

Calculation of the costs of a project is usually straightforward. The only data required is the set of investments required by the project and the selection of an appropriate discount factor to account for the inter-temporal nature of the investment flow. Estimation of benefits is a much more complex task. One of the most common methodologies used to estimate a program's benefits is a surplus approach (for example, used by MCC in their evaluation of road rehabilitation projects in Armenia and Burkina Faso). Figure 2 illustrates the basic idea behind this calculation.

FIGURE 2 THE SURPLUS APPROACH

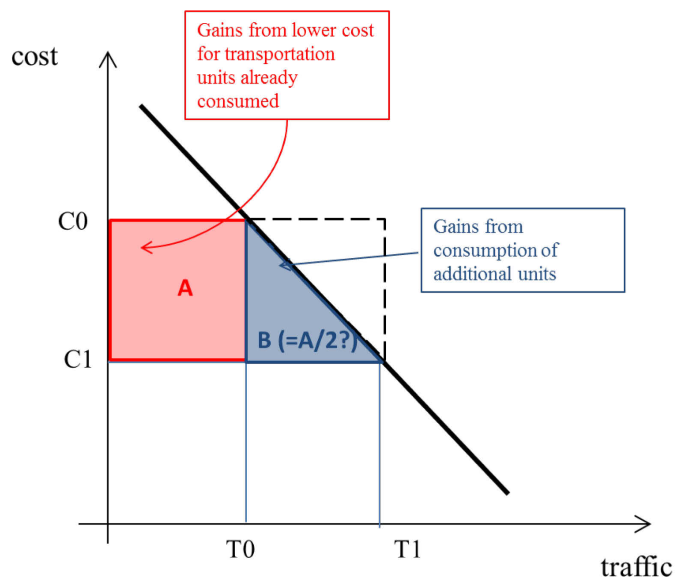


Figure 2 shows the demand curve for transportation, which relates how many "units of transportation" would be "consumed" with different unit-prices of transportation. With an initial cost C0 of transportation, T0 units are consumed. If the transportation price drops from C0 to C1, the accompanying shift in traffic would be T1-T0. Then the surplus generated by the project would be comprised of two areas. The first area is A (in red) and represents the gains from existing traffic (i.e. each of the units already consumed valued at the price differential). The second area is B (in blue) and represents the gains from new "generated" traffic: each of the additional units of traffic that were not consumed before the project (and are consumed after the project) appraised at the unit value determined by the demand schedule.

The implementation of this methodology requires collecting (or assuming) data on traffic for the rehabilitated road. This data allows us to determine the characteristics of the vehicle fleet that usually travels that road. Each type of vehicle (e.g. truck, automobile, motorcycle, etc.) is assumed to have a certain type of motor and tires. Fuel consumption and occupancy are also assumed for each type. The reduction in travel cost (the difference between C0 and C1) is then estimated using engineering models (such as the Highway Development and Management Model, or HDM) that are based on parameters for reduced vehicle depreciation rates (motor and tires), decreases in fuel consumption, and time savings (i.e. travel time reductions multiplied by average hourly wages). All of these components provide the cost reduction per travel unit.

We can readily apply these cost reductions to existing traffic levels to estimate the red square (A) in Figure 2. However, it is more difficult to estimate the blue triangle (B). For this, we would need to know the level of traffic that would be generated by the road improvement. Some studies estimate this triangle to be half of the red rectangle as a rule-of-thumb; however, this approximation overstates the true impact when travel demand is somewhat elastic and understates it when demand is inelastic. The real size of this triangle is hard to calculate without the demand curve (of which little is typically known).

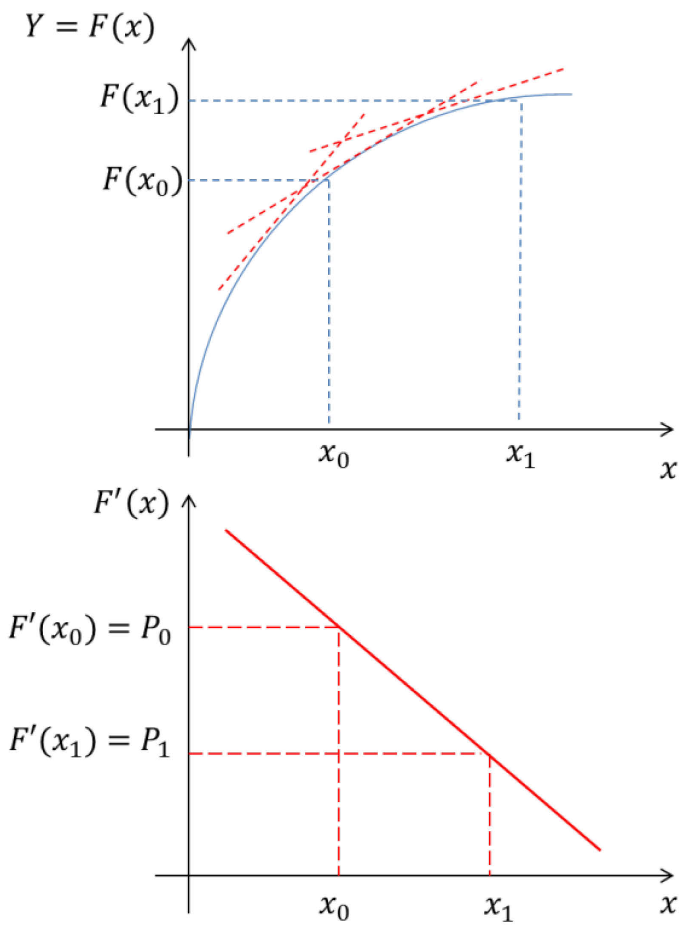
Another approach is to collect traffic data before and after the project is completed; the difference in traffic can be used to estimate the distance T1-T0 and the demand slope can be determined by extrapolating the points (T1, C1) and (T0, C0). While certainly more rigorous than the previous rule-of-thumb, this methodology has disadvantages of its own. Measurements of traffic before and after the project do not necessarily provide an accurate measure of traffic generated by the road itself. For example, if there are any other factors affecting traffic other than road construction (i.e. simplified customs for imported cars, increases in income that allow more families to own cars, etc.); the difference cannot be wholly attributed to the project.

Our approach is somewhat different. Instead of assuming that transportation is consumption good, we treat it as an input in the production function of rural households. In this light, assume that *x* is transportation and $F(x)$ determines the level of production *Y* that corresponds to each level of this input[3]. The demand for factor x is then determined by its marginal productivity (i.e. $F'(x)$); the farmer's willingness to pay for an additional unit of x is precisely what this additional unit would produce.

Figure 3 depicts hypothetical schedules for a production function $F(x)$ and the input demand for *x*. The input demand is determined by the slope of $F(x)$ throughout the range of *x*: $F'(x)$. When the price of factor x is P_0, the farmer demands units of x until $F'(x_0) = P_0$ (analogously, when the price reduces $F'(x_1) = P_1$). Note that the factor demand in the lower panel of Figure 3 is the same as the one in Figure 2 (from which the benefits of the project can be calculated using the surplus approach).

[3] As usual, we assume that the production function is increasing and concave. We normalize the output price to 1 so that $F(x)$ is also a revenue function. However, assuming any other output price does not affect this idea.

FIGURE 3 THE SURPLUS AND PRODUCTION APPROACHES



Rather than estimating the demand curve (or making any assumptions) for transportation, we estimate the difference between $Y_1 = F(x_1)$ and $Y_0 = F(x_0)$. Because the demand curve for x is its marginal productivity, the area under $F'(x)$ between x_0 and x_1 is equivalent to $Y_1 - Y_0$: $\int_0^{x_1} F'(x)dx - \int_0^{x_0} F'(x)dx = F(x_1) - F(x_0)$.

Thus, our methodology relies on directly measuring the change in production (or income, from the different rounds of surveys we collected) derived from the NTH rehabilitation project. This approach has several advantages. First, we can gauge the benefits of the project from observed changes in income, which does not require any assumptions about the input demand function (or the production function). Second, we do not need to rely on assumptions regarding depreciation factors or to measure households' time savings. Third, instead of capturing benefits from traffic flows as in the HDM models (which include foreign companies, large firms in the cities, etc.), we can restrict our analysis to the population of interest: rural households in the NTH's area of influence. Finally, the micro data enables us to capture heterogeneous treatment effects, which allows us to conduct a more accurate beneficiary analysis of the impact of the project.

3. Literature Review of the Evidence

The importance of transportation infrastructure for market expansion and division of labor was recognized as early as 1776 [(Smith, 1776)]. Similarly, early work on economic growth and development included physical infrastructure in the category of social overhead capital and highlighted the need for such capital as the basis for development [(Hirschmann, 1959)]. Recognizing the role of infrastructure in economic development, Smith asserted that social overhead capital "is usually defined as those services without which primary, secondary and tertiary production activities cannot function" [(Hirschmann, 1959)].

Though the empirical literature on the impact of infrastructure only emerged in the mid-1980s, it has grown since then to encompass a large and diverse body of work. Different methodologies have been used, ranging from production functions [(Holtz-Eaakin, 1994); (Garcia-Mila, McGuire, & Porter, 1996)] to cost functions (Nadiri & Mamuneas, 1994); (Morrison & Schwartz, 1996)]. Studies have also used data with various levels of aggregation such as firm-, household-, or village-level microeconomic studies [(Antle, Human capital, infrastructure, and the productivity of Indian rice farmers, 1984); (Ahmed & Hossain, 1990); (Dong, 2000) or state- (Aschauer, Is public expenditure productive? , 1989) and country-level macroeconomic studies (Aschauer, Public Investment and productivity growth in the group of seven. , 1989). Studies have also focused on specific sectors such as manufacturing (Morrison & Schwartz, 1996) or agriculture [(Binswanger, Khandker, & Rosenzweig, 1993); (Paul, Ball, Felthoven, & Nehring, 2001)]. Others have taken a regional approach such as developed (Röller & Waverman, 2001) versus developing countries (Fan, Hazell, & Thorat, 1999). There are also empirical studies that examine the impact of a specific infrastructure such as roads [(Jacoby H. , 2000); (Jacoby & Minten, 2009) ; (Gibson & Rozelle, 2003)] and phones (Norton, 1992)].

(Antle, Infrastructure and aggregate agricultural productivity: International evidence, 1983) pioneered the inclusion of infrastructure in studies explaining differences in agricultural productivity across countries. Using aggregate agricultural production data for 1965 from 47 developing and 19 developed countries, Antle employed a Cobb–Douglas production function that included infrastructure as a production input. Infrastructure was defined as a gross domestic product of the country's transportation and communications industries, measured per square kilometer of land. Not surprisingly, infrastructure appeared to have strong positive impacts on agricultural productivity both in developed and in developing countries. In a similar study on Indian farmers, (Antle, Human capital, infrastructure, and the productivity of Indian rice farmers, 1984) found infrastructure capital to have a systematic effect on farm productivity.[4] A number of econometric problems are associated with this study, however. First, while it is tempting to infer a causal relationship from public capital to output, it is equally likely that the direction of causality goes from output to public capital. Second, common trends in infrastructure and output —that is, the estimated coefficient— may reflect a spurious correlation between output and public capital stock that is

[4]The study included transportation costs proxied by the geometric mean distance to the nearest bus, rail, and postal facilities (in kilometers) in a Cobb–Douglas production function.

driven by a common time trend and not by any underlying relationship between the two variables. Third, omitted variables could create a bias in the coefficient estimates.

(van de Walle D. , 2009) proposed a guideline for impact evaluations of rural roads projects in order to shed some light on the various evaluation design methods. Specifically, she summarized findings from rural road impact evaluation papers that dealt expressly with selection. The use of panel data in impact evaluations is strongly supported by the literature. (Khandker, Bakht, & Koolwal, 2009) used a panel fixed effect estimation with controls for initial conditions to assess a project in Bangladesh and found that investment in roads reduced poverty by raising agricultural production, wages, and output prices and lowering input and transport costs. They also found improvements in schooling outcomes for children, as well as higher impacts for poorer households. Additionally, (Jalan & Ravallion, 2002) used a dynamic consumption growth model with Chinese household panel data covering five years and found that road density has a positive impact on subsequent consumption growth. (Dercon, Gilligan, Hoddinott, & Woldehanna, 2009) also used panel data to assess the impacts of changes in road access on consumption in Ethiopia.

There are many studies that use double difference with a propensity score matching approach. (Lokshin & Yemtsov, 2005) studied the impact of road rehabilitation in Georgia and found that impacts vary between poor and non-poor households. Additionally, in the aggregate, opportunities for off-farm and female wage employment were significantly increased in treatment versus control villages. However, further disaggregation indicates that off-farm employment improved solely for non-poor households, while female wage employment increased for poor women only.

More recently, (Mu & van de Walle, 2011) tested whether the impacts of rural road improvements were consistent with the development of transport-induced local markets in Vietnam between 1997 and 2001. Using double differences and propensity score matching on pre-intervention covariates, they found that the expected outcomes took some time to respond. Their study finds a switch from agricultural to non-agricultural activities, as well as impacts on primary school completion rates. They also find considerable impact heterogeneity, with poorer communities seeing higher impacts.

Another important aspect in the impact evaluation of roads is the time in which one can expect outcomes to change. On this issue, the literature provides little to no guidance. As discussed in (van de Walle D. , 2009), an evaluation should allow for sufficient time for impacts to manifest and acknowledge the differences between short term impacts and medium/long term impacts. The best strategy in this case seems to be a mix of timing in the evaluation of roads to allow comparing how initial impacts evolve and how other impacts might arise in the longer term. The periodicity of this strategy might reduce sample attrition and provide a better way to control for unrelated shocks and contamination.

The majority of studies recognize that investment in roads has a strong impact on rural incomes, especially for smallholders. However, the literature has not been completely successful in assessing the benefits and costs of alternative investments or the causality of relationships that generate higher rural incomes due to better infrastructure services. This gap in knowledge regarding causal relationships between investment in infrastructure services, on one hand, and increases in income-generating opportunities and welfare benefits for rural populations, on the other, limits the scope for specific policy recommendations.

3.1.Evidence Gaps Filled by the Current Evaluation

This impact evaluation proposes a way to quantify the extent to which road benefits accrue to rural population, as well as the incidence of these benefits among the poor. This evaluation will bring much needed hard evidence to the literature on the effects of this type of infrastructure project. As discussed previously, the current literature is plagued by endogeneity problems and omitted variables biases. For example, if relatively well-off areas with higher levels of non-farm activity attract more infrastructure projects, then the positive correlation we observe between infrastructure and income would not be causal due to the endogenous placement of the project. In the same way, if infrastructure is placed in areas with higher unobserved productivity levels, any estimated effect would be biased because of these unobserved factors.

By using unique data and combining two methods (discussed in Section 4.3) specifically designed to measure the impacts of this infrastructure project, we account for time invariant characteristics that could explain road placement. In addition, the availability of multiple rounds of data for the beneficiary households permits us to track the evolution of benefits at different stages of market accessibility, which sheds some light on the sustainability and cost effectiveness of this kind of infrastructure project.

4. Evaluation Design

4.1.Evaluation Type

The benefits of the connectivity project will be measured using a rigorous impact evaluation methodology. An impact evaluation is a study that measures changes in outcomes affecting wellbeing that can be attributed to a specific intervention. Impact evaluations require a credible and rigorously defined counterfactual that estimates what would have happened to the beneficiaries in the absence of the project. Estimated impacts, when contrasted with total related costs, provide an assessment of the intervention’s cost effectiveness.

4.2.Evaluation Questions

4.2.1. Country-specific and International Policy Relevance of Evaluation

Roads are a basic input for all economic sectors and have a large potential impact on development. Transportation determines an important portion of transaction costs, which have several economy-wide implications. High transaction costs hinder competition, arbitrage between markets, market integration, labor mobility, and the creation of value chains.

In rural areas, roads help households to integrate into markets, which allow them to increase their monetary income, access better inputs, coordinate with other actors along the value chain, and purchase goods to expand their consumption basket. In El Salvador, about half of the road infrastructure is unpaved[5]. This limits transportation ability and inhibits economic development in El Salvador. The situation for rural Salvadoran households is even more critical because they are usually located in remote areas where dirt roads become impassable during the rainy season (which accounts for nearly half of the year).

4.2.2. Key Outcomes Linked to Program Logic

It is expected that this project will reduce transportation costs and enable households to extend their labor activities and diversify their income sources.

The first questions are related to the project outputs (direct products of the program, such as segments of the NTH constructed, etc.) as opposed to outcomes or impacts (changes access to markets, changes in income).However, the answers to these questions are qualitative in nature.

- Was the NTH implemented according to plan?
- Did the NTH reach the originally intended beneficiaries? Did it reach unintended population or sectors of the economy?

The connectivity project initially consisted of a network of connecting roads (NCR) and the Northern Transnational Highway but because of the significant increase in construction costs and the existence of other interventions in connecting roads the project only focused on the Northern Transnational Highway. Table 1 describes the evolution of the project components as well as the current investment. As a result of these changes the current impact evaluation design focuses only on the evaluation of the Northern Transnational Highway.

On the beneficiary question, it is expected that since the network of roads was not built, the number of beneficiaries could be smaller. However, given the nature of the NTH investment and the geographical area that it spans, it is likely that the number of beneficiaries did not change considerably since the "beneficiary" definition was changed from "individuals living in a radius of 2 Km of the roads" to "individuals living 30 minutes from the NTH or within 5 Km" when the NCR was dropped. At closeout of the compact, the estimated beneficiaries of the NTH are 533,667 individual that live within 5 km of the NTH [(MCC, 2012)].

[5] The World Bank's World Development Indicators estimate that the paved share of the total road network in El Salvador was 54.1% in 2009 and 46.9% in 2010.

TABLE 1 INITIAL CONNECTIVITY PROJECT COMPONENTS

Components	Expected number of beneficiaries	Initial Expected Effects	Initial Estimated Investment	Current Estimated Investment
Network of Connecting Roads (NCR)	600,000 individuals (i.e. population 2km from either side of the road). This accounts for 70% of the total population of the Northern Zone	Reduced transportation costs Time savings Increased prices for land along the roads	$ 94 million (NCR)	Was not built
Northern Transnational Highway (NTH)		Increased income by at least 5-6% for those households within 2 km of the NTH	$140 million	$269 million

Additionally, the construction of the NTH was split in sections. The timeline of construction of the different sections changed, as described in Table 2; and one segment 7 that was not initially planned to be constructed was added later to the timeline of the compact.

TABLE 2 NTH CONSTRUCTION TIMELINE: INITIAL AND ACTUAL

	INITIAL		ACTUAL	
Section	**Scheduled Start Date**	**Scheduled End Date**	**Start Date**	**End Date**
T1	Will not be constructed		Will not be constructed	
T2	May-09	Oct-10	May-09	Sep-10
T3	Dec-09	Oct-11	Dec-09	Aug-12
T4	Jan-10	Jan-12	Jan-10	Nov-12
T5	Oct-09	Aug-11	Oct-09	Aug-12
T6	Oct-09	Aug-11	Oct-09	Aug-12
T7	Will not be constructed		Sep-11	Sep-12
T8	Constructed		Constructed	

a. Impact on Agricultural Transportation Costs and Incomes

- Does access to the improved NTH reduce agricultural transportation cost? Specifically, does it affect input transportation costs and freight costs?
- Does access to the improved NTH improve market participation by increasing the likelihood of going to the market and/or the volume sold in the market?
- Does access to the improved NTH increase income from agricultural sources?

The first impact we expect is an increase in income resulting from reduced transportation costs for on-farm activities. To illustrate this case, assume a household that purchases *x* units of farm inputs, with a unit cost of *c*. This unit cost includes the price of the direct cost of the input as well as a transportation cost (e.g. the farmer has to travel to an input market and bring the input back to the farm). Its agricultural production function is given by $Q = F(x)$. If the household decides to sell its production in the market, it receives a price of *p* but incurs an output transportation cost of *t*. The household can also decide to self-consume some (or even all) of its production instead of selling it; in this case, we denote self-consumption as *q* so that the households' sales volume is $F(x) - q$. Assume that the household's utility function is quasi-linear in income so that its maximization problem is:

$$Max\ \ U(x, q|p, t, c) = (p - t)(F(x) - q) - cx + V(q)$$

where $F(x)$ and $V(x)$ are increasing and concave functions (i.e. $F'(.) > 0$, $F''(.) < 0$, $V'(.) > 0$ and $V''(.) < 0$). The optimal values x^* and q^*satisfy:

$$F'(x^*) = \frac{c}{(p - t)}$$

$$U'(q^*) = (p - t)$$

We posit that the project should have two effects on households' agricultural activities. The first is a potential reduction in *c*. If households have better connectivity, they can have better access to input markets (which reduces the transportation portion of c). In such a case, it can be shown that $dx\ /\ dc = 1\ /\ ((P - t)\ F''\ (x)\ ) < 0$. Thus, reductions in c should lead to both increases in input demand and higher outputs. Second (and more importantly), the project also leads to reductions in output transportation costs.

$$\frac{dF(x)}{dt} = \frac{\mathrm{F}'(\mathrm{x})}{(p - t)F''(\mathrm{x})} < 0$$

$$\frac{dq}{dt} = -\frac{1}{U''(q)} > 0$$

In this line, we expect that reductions in transportation costs will lead to enhanced profitability of households' sales. Thus the household will be both induced to increase its production and to sell more (through more active market participation and reductions in self-consumption).

b. Impact on Off-farm Income, Income Diversification and Time Allocation

- Does access to the improved NTH increase the availability of non-farm employment?
- Does access to the improved NTH promote the creation of non-farm enterprises?
- Does access to the improved NTH increase income from non-farm sources?
- How does access to the improved NTH affect the time allocation across labor and leisure activities? How does it change the labor allocation between farm and non-farm activities?

More broadly speaking, the intervention is expected affect non-agricultural activities as well as agricultural ones. For example, better access to roads could allow household members to commute more readily to non-farm jobs, enlarge markets for their non-farm products, and generate more opportunities for income diversification.

The analysis framework for this possibility is outlined below. Income of a rural household *i* can be expressed as the sum of incomes that the household receives for *J* different activities (e.g. farm and non-farm activities):

$$Y_i = \sum_{j=1}^{J} y_{ij}$$

where Y_i represents total income of the *i*-th household and y_{ij} represents its income from activity *j*. Each activity-specific income y_{jj} can be decomposed into two components: hours worked (l_{ij}) and the hourly return (y_{ij}/l_{ij}) of the *i*-th activity, such that:

$$Y_i = \sum_{j=1}^{J} l_{ij}\left(\frac{y_{ij}}{l_{ij}}\right)$$

Finally, the number of hours spent on activity *j* can be expressed as the product of the total hours worked (L) and the share of time allocated to activity *I* (Cl_i). Therefore, we can now express the total income as:

$$Y_i = L\sum_{j=1}^{J} Cl_{ij}\left(\frac{y_{ij}}{l_{ij}}\right)$$

Our hypothesis is that access to the NTH (and the subsequent reduction in transaction costs) will result in a change of income for rural households through an increase in the demand for rural products and a change in prices for both farm and non-farm products. This change in income (ΔY), which is obtained by a household because of better access to transportation, can then decompose in the following way:

$$\begin{aligned}\Delta Y_i \quad = \quad & \Delta L \sum_i Cl_i \frac{y_i}{l_i} + L\sum_j \left[\Delta Cl_{ij}\frac{y_{ij}}{l_{ij}} + Cl_{ij}\Delta\left(\frac{y_{ij}}{l_{ij}}\right) + \Delta Cl_{ij}\Delta\left(\frac{y_{ij}}{l_{ij}}\right)\right] + \\ & +\Delta L \sum_j \left[\Delta Cl_{ij}\frac{y_{ij}}{l_{ij}} + Cl_{ij}\Delta\left(\frac{y_{ij}}{l_{ij}}\right) + \Delta Cl_{ij}\Delta\left(\frac{y_{ij}}{l_{ij}}\right)\right]\end{aligned}$$

This equation decomposes changes in income into three building blocks. The first block captures the impact of changes in the total number of hours worked, *keeping labor allocation constant* (i.e. $\Delta L\left(\sum_i Cl_i \frac{y_i}{l_i}\right)$. The second block captures the impact of labor allocation, *keeping total hours of labor fixed.* This effect has three subcomponents: (a) change in shares of hours spent on each activity, *fixing the monetary return of each activity*, i.e. $\Delta Cl_{ij}\frac{y_{ij}}{l_{ij}}$; (b) changes in the monetary return of each activity, *fixing the shares of each activity, i.e.* $Cl_{ij}\Delta\left(\frac{y_{ij}}{l_{ij}}\right)$, and (c) the interaction between these two sub-components.

The last component of the equation represents the interaction between the first two components (i.e. changes in the total hours worked and in labor allocation).

This approach allows us to determine the extent to which three different factors affect rural households' incomes: total labor supply, shares of time allocated to each activity, and differential returns between agricultural and non-agricultural labor.

c. Impact on Human capital, Consumption and Land

- Does access to the improved NTH increase the use of public services? Specifically, health and education services?

NTH not only provides better access to input and output markets; it also reduces travel times to schools, health centers, and public services in general. This can have important effects on household welfare and human capital accumulation.

- If access to the improved NTH increases income, how does this reflect in the consumption patterns of households? Is there an effect on food consumption and/or non-food consumption? How are they different?

The project can also have an important effect on consumption patterns. First, there is an income effect that can relax budget constraints (i.e. more money for purchases). Second, market access also provides the opportunity to purchase goods that might not have been available before. Third, as mentioned previously, we expect that households will reduce their self-consumption of agricultural products (which usually entails few and relatively non-perishable products) and increase their market purchases. All of these factors can have important consequences for households' food consumption and can also potentially affect their non-food purchases.

- Does access to the improved NTH increase land investments and land values in the northern zone?

The project might increase land values through two potential channels. The first is land as an agricultural input (x) in the model outlined above. As such, an increase for this production factor can improve land values. Second, better connectivity can also lead to changes in land use due to the expansion of sub- and peri-urban areas (with further increases in land demand).

d. Gender and Social Exclusion

- Do the effects on farm and off-farm activities and income differ by gender or by expenditure levels (initial conditions)?
- How are the effects on health and education access different for men and women? How are they different for the extremely poor versus the relatively poor?
- What factors (use of time behavior, sources of income, etc.) might explain the impact (or lack of impact) in a specific subpopulation?

We will also try to identify differentiated gender effects through a different set of variables: for example, total changes in income, changes in hours worked, changes in hours worked in non-paid activities, changes in hours worked in non-farm activities, hours spent on childcare and household chores, etc. In addition we will explore possible differential effects across the poverty scale by using the poverty status at baseline to classify the population. Differential effect by initial poverty condition are important as they would show if the greater, smaller or lack of impact in one subpopulation might be explain by the level of poverty in which they started; in essence, some people might be too poor to be able to experience improvements in the welfare outcomes targeted by the project, for example landless households and people out of the labor force.

e. Macro-regional impacts

- What is the impact of the NTH in the entire economy of the northern zone of El Salvador? And across other regions of the country?

We will use dynamic regional computable general equilibrium (CGE) model for El Salvador that incorporates regional disaggregated sectors for agriculture (North, South, and central) to measure the effects of the changes in accessibility on the northern region with respect to all other regions of the country. Using this model, we will simulate the impacts of NTH investment (as evidenced by the impact evaluation) and measure the effects across regions on i) productivity increases in the agricultural sector, ii) increases in GDP, iii) increase in exports of tradable agricultural commodities, and iv) increases in employment and household incomes and their consequent distributional effects.

In summary, we present Table 3 with the expected effects for the outcomes discussed above.

TABLE 3 EXPECTED EFFECTS FROM THE CONNECTIVITY PROJECTS OVER MAJOR OUTCOMES

Indicator	Expected Effects	Gender	Data Sources
Income	Positive	Larger effects on women	Household survey
Consumption patterns	Positive	Larger effects on women	Household survey
Number of hours worked	Positive	Larger effects on women	Household survey
Numbers of hours in leisure	Not Clear	Negative effects on women	Household survey
Number on hours in non-farm productive activities	Positive	Larger effects on women	Household survey
Number of hours in child care	No effect	Potentially Negative on women	Household survey
Land Values	Positive	No differentiated effects	Household and Community survey
Traveling time to markets	Negative	No differentiated effects	Household and Community survey
Traveling times to public services	Negative	Larger effects on women	Household and Community survey
Transaction costs for agricultural producers	Negative	No clear differentiated effects	Household and Community survey
Specialization of production	Positive	No clear differentiated effects	Household survey
Migration and Remittances	Positive	No clear differentiated effects	Household survey

4.3. Methodology

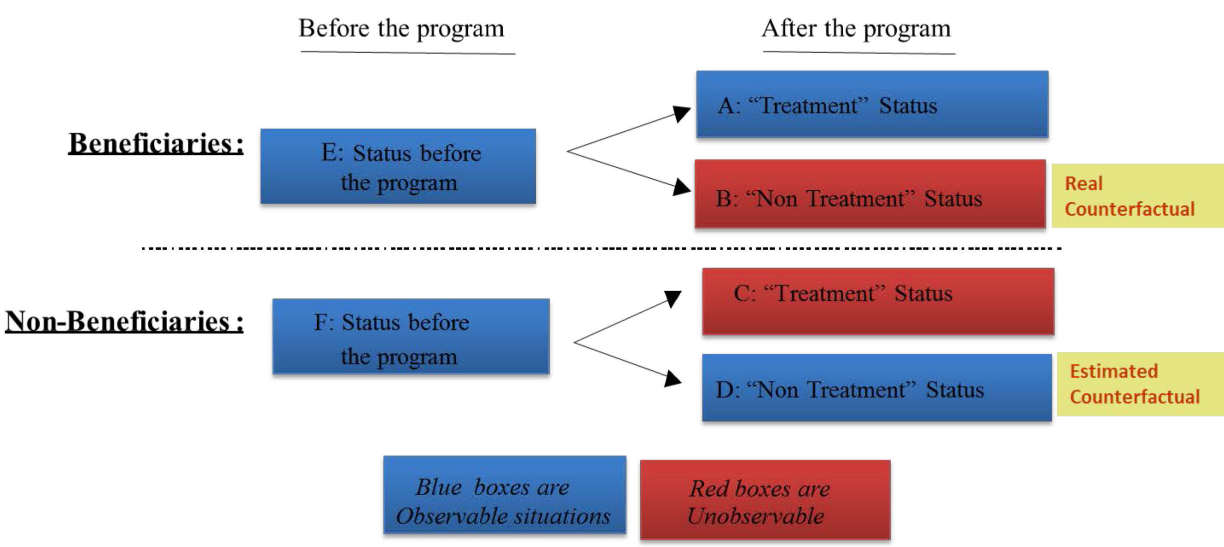
Our evaluation will rely on both a micro and a macro perspective. The former component will look at the household data to analyze more detailed impacts of connectivity on the population's welfare in Northern El Salvador. In this line, we will apply two different quasi-experimental impact evaluation techniques. The first approach is to exploit regression discontinuities. As the NTH is built in different segments, this creates adjacent areas that are differentially exposed to road rehabilitation. Thus, we can compare households in these neighboring treatment and control areas to capture the effect of the project. The second approach exploits the continuous variation in travel times to relevant locations (i.e. markets) resulting from the NTH. In this line, rather than exploiting the dichotomous exposure to the treatment, we can compare households that have experienced larger or smaller variations in transportation time.

The macro component aims to understand the country-wide effects of these road infrastructure investments. For this purpose, we will use a Computable General Equilibrium (CGE) model developed by IFPRI specifically for El Salvador. This model will provide insights about the impact that reductions in transportation times in the northern region can have on national value chains and their distributional consequences.

4.3.1. Micro Component of Impact Evaluation

By conducting an impact evaluation, we intend to quantitatively estimate the change in the population's well-being at the micro level as a result of the program. Thus, we aim to compare the population's well-being once the program has been executed relative to the population's well-being had the program not been implemented. In other words, the basic principle that guides our approach is the comparison between situations with the program and without the program, also known as "treatment effect". This approach is illustrated in Figure 4. In an impact assessment, we would like to compare changes in those who were treated vis-à-vis what would have happened had these treatment not been provided (i.e. boxes A and B). This is opposed to a mere comparison of the situation after the program and before the program (i.e. comparing A to E or the difference between participants and non-participants (A to D)). Unfortunately, it is not possible to observe state B.

FIGURE 4 POSSIBLE SITUATIONS FOR TREATED AND CONTROL HOUSEHOLDS



To address this problem of unobservable situations, we identify a control group (D) that is as similar as possible to the treatment group, so that observations of D are a close approximation of B. Ideally, we would get valid counterfactuals from an experimental approach in which households are randomly assigned to treatment and control groups. Random assignment ensures that the distributions of characteristics (both observable and unobservable) among both groups are statistically indistinguishable. In our specific program – and, more generally, in the provision of infrastructure services – random assignment of such services among households is not feasible since it could conflict with the construction of the NTH. Given that full randomization is not feasible, we implement non-experimental assessment methods. However, it is important to note that none of these methods offers a perfect solution (Ravallion, 2007)

We use a combination of methods to obtain valid inferences about the various household-level impacts of connectivity. These methods include carefully selecting our analytical samples of recipients and non-recipients to be as similar as possible in terms of their observable characteristics prior to the program. The framework serving as a guideline for our empirical analysis is the Roy-Rubin model [(Roy, 1951) (Rubin, Estimating Causal Effects to Treatments in Randomized and Nonrandomized Studies, 1974), (Rubin, Assignment to Treatment Group on the Basis of a Covariate, 1977), (Rubin, Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies, 1979), (Rosenbaum & Rubin, 1983)].

"Roads are clearly not randomly placed, and it is highly likely that the factors that led to the road placement will also affect outcomes" [(van de Walle & Cratty, 2002)]. Thus, since an experimental design is not feasible, we will implement two non-experimental designs.

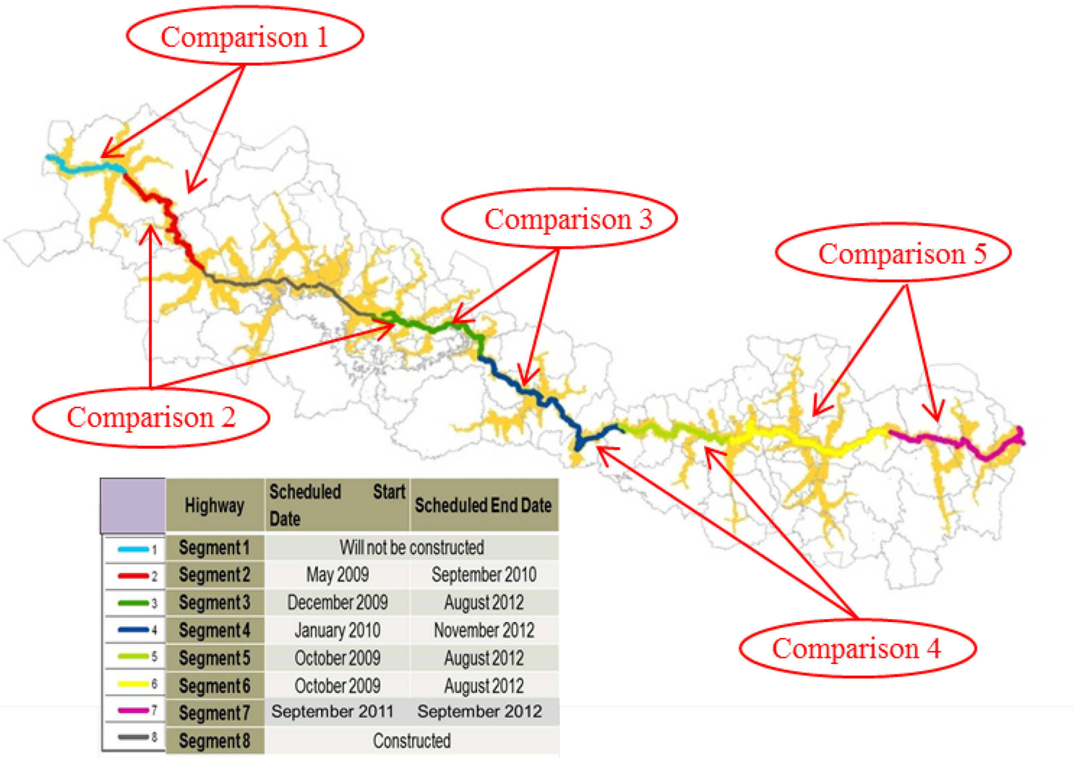
a. The Regression Discontinuity (RD) Approach

The improvement process of the NTH was scheduled to take place between May 2009 and November 2012 and involved eight construction segments. These segments were mostly determined by cost effectiveness concerns (e.g. the presence of natural barriers, engineering factors, etc.). The idea of the RD approach is to exploit these discontinuities, take advantage of the roll-out of the construction of the NTH, and compare households in adjacent segments.

This methodology assumes that households do not self-select into either side of the segment boundaries. Thus, the households on both sides are essentially comparable; they just happen to be divided by an engineering discontinuity that determines the timing in which they benefit from the NTH construction. In this line, the segments can be used as a quasi-random assignment of households into treatment and control groups over time.

While considerably more complicated in practice (because of the multiplicity of segments and implementation dates), this is the basic idea of the RD design. Figure 5 shows the scheduled dates of the construction of the different segments (*tramos*) and the comparison groups generated by this timeline. For example, Segment 2 (scheduled for construction between May 2009 and September 2010) lies to the right of Segment 1 (which was not constructed). These segments comprise the first comparison group. We can estimate the intervention's short- and long-term effects by assessing the differences between Segments 1 and 2 in several rounds of the survey (i.e. one, two, three, or four years after the intervention, using the 2010, 2011, 2012, and 2013 rounds of the survey). A similar approach allows us to determine four other comparison groups for the RD estimation.

FIGURE 5 ROLL-OUT OF NTH AND COMPARISON GROUPS FOR RD

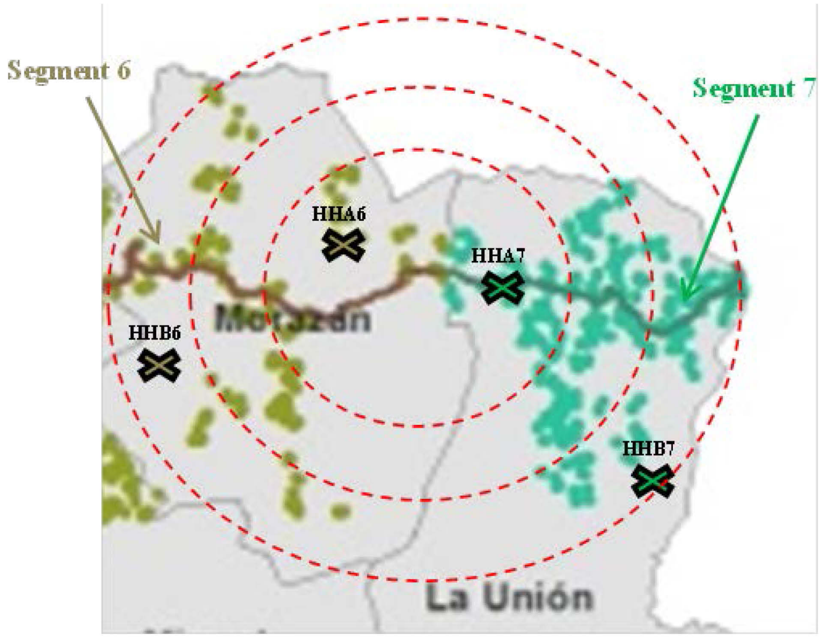


We adjust comparability between comparison groups using two further procedures. First, we restrict our sample to those households who live 30 minutes away from any planned segment of the NTH at most. This restricts the evaluation to households that would be most likely to benefit directly from road improvement.

Second, this approach will assume that adjacent segments are somewhat similar and comparable. While this is intuitively plausible in the borders of a segment (i.e. these households are very similar and just happen to live on either side of the discontinuity), it may not hold for all households in the comparison groups. This will be increasingly unlikely for households that are farther away from the discontinuity. For this reason, we also plan to adjust the comparability between households in the treatment and control segments using buffers.

For example, consider two pairs of households A and B in segments 6 (HHA6 and HHB6) and 7 (HHA7 and HHB7) in Figure 6. It is likely that households HHA6 and HHA7 are more comparable than HHB6 and HHB7. Thus, we can establish a smaller buffer to guarantee a tighter comparison among groups (and exclude less comparable households HHB6 and HHB7). However, excluding observations that are farther away also reduces the efficiency of the estimator. In light of this, we propose to include estimations with different buffer sizes and compare these results.

FIGURE 6 COMPARISON BUFFERS FOR RD



b. The Continuous Approach

The goal of the previous approach is to capture the difference in outcomes between households in areas with and without roads. While the simplicity of this approach is appealing, we need to determine whether each household has been treated or not, which is not easy in this context. For example, consider two households (Figure 7) that go to a nearby market (blue "X" on the bottom of the graph) to sell their harvests. Household A lives in Segment 6 (where NTH was constructed between October 2009 and August 2012) and Household B lives in Segment 7 (where NTH was only constructed in the last year of the project). The dashed line indicates the path that each household would travel to reach target X.

The RD approach would have assigned A to the treatment and B to the control group for the period in which Segment 7 was still not constructed. However, even before the construction of Segment 7, Household B is not totally unaffected by the NTH. While most of its route to the market will go through the unimproved (green) section of the road (and thus see no changes in its travel time), B's route does include a small section of the improved road (brown); it will therefore save some time from improvement of the NTH. Household A lives in a segment with an improved section of the road and will thus spend more time on the NTH than Household B.

Assume that transportation time to the market is inversely related to households' agricultural income. This hypothetical relationship is the depicted by the function $Y = F(travel\ time)$ in Figure 8. The figure shows that Households' A reduction in travel time will translate into larger increases in its agricultural income. While more modest, Household B will also experience some improvement in income.

All in all, the continuous approach does not assign households to a treatment and control group but instead exploits variations in the degree to which NTH reduces households' transportation time and the extent to which these reductions translate into enhanced welfare outcomes.

FIGURE 7 HYPOTHETICAL TRANSPORTATION ROUTES OF HOUSEHOLDS IN ADJACENT SEGMENTS

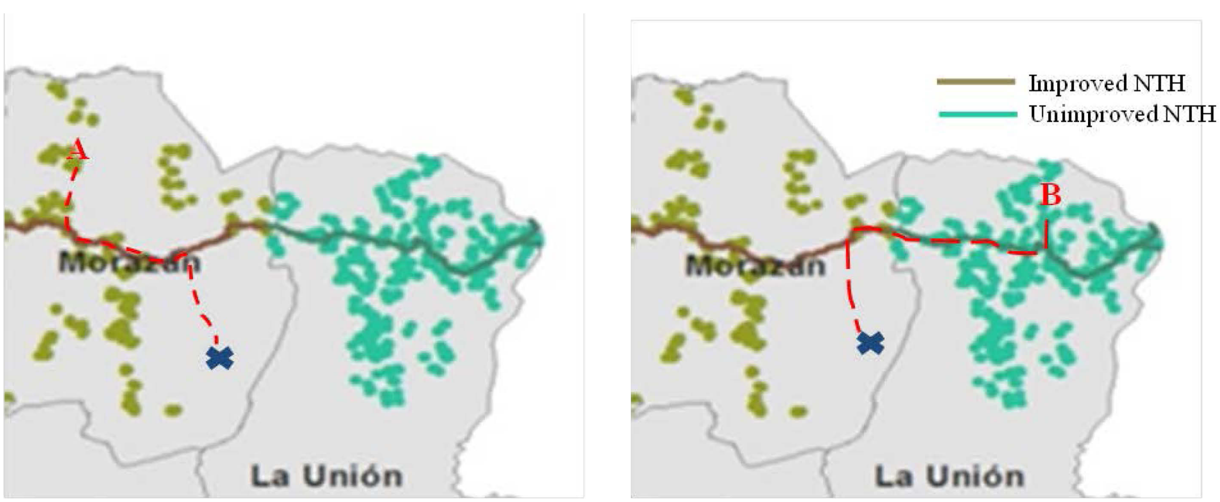
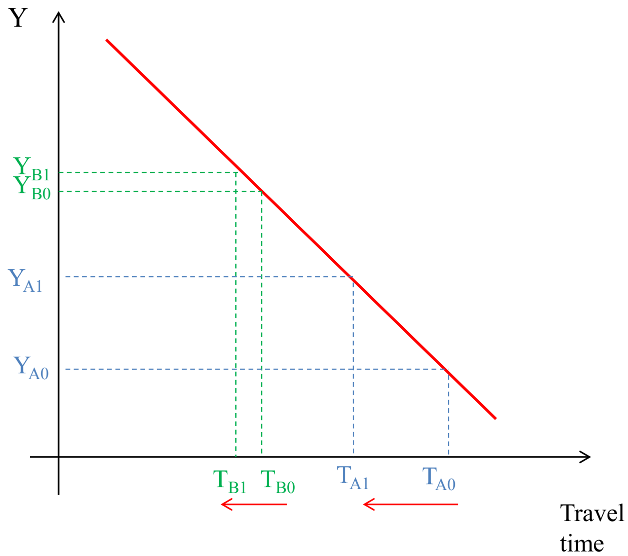


FIGURE 8 HOUSEHOLDS' BENEFITS FROM THE PROJECT



The continuous approach is conducted in several different stages:

1. First, we need to determine a set of relevant destinations in a given area. While agricultural markets are not the only potential targets, they represent an obvious candidate due to their economic relevance for rural households.
2. Once the relevant targets have been defined, we need to estimate each household's travel time. For this purpose, we will apply raster analysis. This method calculates the shortest time from any household to relevant local or regional markets. It considers the availability of different road surfaces (e.g. paved road, dirt road, no road, etc.) and their respective impedance factors, which reflects traveling speeds on different quality roads and on variously sloped terrains through which the road passes. This procedure allows us calculate travel times to different markets and determine for each household the market with the shortest travel time. Using road data in the baseline, this procedure provides us with an initial "optimal" travel time for each household.
3. Third, we re-estimate households' travel time after the implementation of the project. For each period, we alter the impedance factor of the segment of the NTH that has been improved (which captures road enhancement and higher speeds of transit) and re-estimate the "optimal" travel time under these new conditions.
4. Fourth, we use the variation in travel time experienced by each household in a regression setting to determine its effect on households' income and other welfare indicators.

All in all, while the RD approach allows for simpler and more intuitive calculations of the impact of the project, the continuous approach exploits a more detailed source of variability in transportation time resulting from the NTH.

4.3.2. General Equilibrium Component of Impact Evaluation

We also want to capture the indirect effects of the improvement on households' accessibility. For this, we propose to use the IFPRI General Equilibrium Model (CGE) for El Salvador.

CGEs are a good tool to measure the impacts of improvements in infrastructure on an entire economy because they can track the changes in productivity and decreases in transportation costs caused by new roads, as well the effects of these changes on the rest of the economy. IFPRI has developed a dynamic regional computable general equilibrium (CGE) model for El Salvador that incorporates regional disaggregated sectors for agriculture (North, South, and Central); the base year of the Social Accounting Matrix is 2005. The advantage of having a sub-regional model is that it allows us to measure the effects of the changes in accessibility on the northern region with respect to all other regions of the country. The model is useful as a development tool because it allows us to determine the effects of regional investments intended to reduce regional poverty, as well as to explore policy options to deal with a number of macro and balance-of-payments issues that may arise from such investments. The regional nature of the model also permits us to examine the impact of sectorial development policies, particularly those focused on agriculture. It is important to note that in the regional SAM for all agricultural sectors, regional activities that produce the same commodity are disaggregated by the four regions and then combined into one national commodity.

Using this model, we simulate the NTH investment targeted to the northern region and measure the effects across regions on i) productivity increases in the agricultural sector, ii) increases in GDP, iii) increase in exports of tradable agricultural commodities, and iv) increases in employment and household incomes and their consequent distributional effects.

4.4. Population Being Studied

The population being analyzed consists of the people living within a 30 minute radius of the NTH. This region (the Northern Zone) contains one-half of El Salvador's poorest municipalities and suffered more damage from the country's internal conflict during the 1980s than any other region. Economic and social indicators in the Northern Zone are currently worse than the national averages.

4.5. Power Analysis and Sample Size Requirements

It is of vital importance in impact evaluation studies to address issues of power and sample size at the design stage of the study. With that in mind, in the initial design we used EHPM (Encuesta de Hogares de Propósitos Múltiples) and census data to construct a wide range of plausible scenarios, and we considered cases of discrete and continuous treatment variables. The initial power analysis was conducted under the assumptions that the minimum detectable impact/effect should be a 20% increase in household income, with 80% power at the 5% confidence level. It concluded that our survey needed a sample size of at least 3,775 observations to be able to detect the effect of the road improvement program on rural (total household) income/expenditure, while also being representative of the highway's different segments. Due to the high intra-cluster correlations observed in variables such as non-agricultural salaried income or time allocated to non-agricultural non-wage labor, the power needed to detect differences in such variables is lower.

In “Annex 1: Original Sample Design” we describe the original sample design, based on the EHPM calculations. Below we provided updated power calculations using data from the four survey waves available to date (2009-2012) and taking into account the changes due to the sample frame and the expected benefits. In summary, the results across methodologies show that the survey is powered to detect only large effects using the discrete treatment methodology (RD), from 15 to 29 percent changes in expenditures depending on the segments being compared, which is above of the 6 percent income increase revised target expected by MCC. Using the continuous treatment methodology it is powered to detect an effect equal to 9.5 percent of baseline income per hour of time gained. We note that even though this seems a large gain, the impact estimate using this methodology is a combination of the differences in means and the detected “structural” relationship between the outcome and the travel time variables. This implies that we will be able to detect small impacts using this sample using the continuous treatment. For example, we could detect an impact as low[6] as 6% of the baseline expenditure if the average reduction in travel times is 38 minutes (which is considerably large), depending on the different gains household might have (since these gains are in the order of minutes)[7].

4.5.1. Updated Power Calculations

In this section we review the initial sample size recommendations and updated the parameters used for power analysis using the baseline data, to emulate the calculations based on the EHPM and then proceed to account for the overtime variation in outcomes given that we have four rounds of data from the impact evaluation survey. We present the calculations using a discrete treatment assignment, meaning that households are designated as treatment or control depending in what segment of the NTH they live; and a continuous treatment, where the treatment designates the intensity of the expected benefit, in this case the reduction in travelling times to access a market.

We assumed a clustered, quasi-randomized evaluation design with treatments administered at the cluster level and with data collection before and after initiation of the treatments. The purpose of the sample size estimates or power analysis is to determine the minimum detectable impact (Δ) for a given number of sampled clusters (*g*) and households per cluster (*m*) in each treatment condition for the evaluation sample.[8] If the impact of the treatment is at least as large as Δ, we will be able to detect it at the 95 percent confidence level with the assurance that at least 80 percent of the time that the null hypothesis of no impact is false (i.e. there is an impact), we would reject this null hypothesis if we have a sample of

[6] The variance of the impact estimate is a function of three random variables: the estimated parameter, the mean at baseline, and the mean at follow-up. This variance can be approximated by the delta method, the formula for this variance is complicated; so we just note that the impact of the estimate will depend on the inter-period correlation of the travel times and the variance of the estimate.

[7] The power calculations for the discrete treatment are done using the baseline data. We have spent time explaining the biases in cross-sectional estimates of the relation between access and income. However, given that a priori one does not have a reliable estimate of this parameter at the sample design stage, we opted to replicate as best a possible what one would do given the available data at the sample design stage.
One of the advantages of this research is that it could serve as a benchmark which future studies could use to conduct ex-ante power analysis.

[8] In addition to *g* and *m*, the minimum detectable impact, Δ, is a function of the variance of the outcome variable, its intra-cluster correlation, and the area of influence of the highway being evaluated. See formulae below

total size *2mg*. If the treatment impact is less than Δ, we are less likely to detect it, although detection is still possible.

a. Sample Conditions for Discrete Treatment

With the discrete treatment, the impact evaluation estimate is equivalent to a difference-in-difference estimator that compares households living in segments across time and space. This methodology requires repeated household observations. Power calculations for this type of survey design were based on (Murray, 1998). The main analysis is based on the following three equations:

(1)

$$g = \frac{2\left(1 + (m-1)\widehat{ICC}\right)\left(t_{\alpha/2} + t_{\beta}\right)^2}{m\,\widehat{\Delta}^2}\hat{\sigma}^2{}_y$$

(2)

$$\hat{\sigma}^2{}_y = \frac{mg}{2(1 + (m-1)\widehat{ICC})}\hat{\sigma}^2{}_\Delta$$

(3)

$$\hat{\sigma}^2{}_\Delta = 4\left[\frac{\hat{\sigma}^2{}_m\left(1 - \hat{r}_{yy(m)}\right) + m\hat{\sigma}^2{}_g(1 - \hat{r}_{yy(g)})}{mg}\right]$$

where:

g	:	number of clusters in each condition (treatment/control-cantons)
m	:	number of observations per cluster (households)
$\widehat{ICC}$	:	intra-cluster correlation
α	:	type I error rate
β	:	type II error rate
$\hat{\sigma}^2{}_y$	:	estimated variance of the outcome variable
$\widehat{\Delta}$	:	estimated change
$\hat{\sigma}^2{}_\Delta$	:	estimated variance of the change in the outcome variable
$\hat{r}_{yy(g)}$	:	overtime correlation within groups
$\hat{r}_{yy(m)}$	:	overtime correlation within members

Replacing (2) in (1), we get:

(4)

$$1 = \frac{\left(t_{\alpha/2} + t_{\beta}\right)^2}{\widehat{\Delta}^2} \hat{\sigma}^2{}_{\Delta}$$

Inserting (3) in (4):

(5)

$$1 = \frac{4\left(\hat{\sigma}^2{}_m\left(1 - r_{yy(m)}\right) + m\hat{\sigma}^2{}_g\left(1 - r_{yy(g)}\right)\right)\left(t_{\alpha/2} + t_{\beta}\right)^2}{mg\,\widehat{\Delta}^2}$$

Solving for g:

(6)

$$g = \frac{4\left(\hat{\sigma}^2{}_m\left(1 - r_{yy(m)}\right) + m\hat{\sigma}^2{}_g\left(1 - r_{yy(g)}\right)\right)\left(t_{\alpha/2} + t_{\beta}\right)^2}{m\,\widehat{\Delta}^2}$$

In updating the power calculations, we use the 2009-2010 baseline survey to calculate the intra-cluster correlation and variance needed to calculate the power of the survey; we then follow this with a power analysis focusing on the minimum detectable effects[9] (MDE) for the sample size we have in the survey. The former serves to contrast the initial recommendations based on the EHPM and what the recommendations would be using the baseline survey data. The latter serves as a post-hoc power analysis to revise our expectations regarding the precision and power of any impacts that are detected. These revisions take into account the changes in the sample due to issues of representativeness, errors in the sampling frame, and changes in the timing of the construction of the different segments of the highway. The general purpose is to know what differences we are able to detect in the final impact analysis and to weigh any gains or losses that might have occurred since the initial sample design.

In the following tables, we calculate the sample size needed given the observed mean and variances in the baseline survey and present two scenarios: one using the standard 80% power with a 5% confidence level in Table 4 and another using the standard 80% power with 10% confidence level in Table 5. The results are made under the following parameters, assumptions, and estimations[10]:

Mean Monthly Household Expenditure (2009)	$254
Variance of Mean Household Expenditure	37,917
Minimum Detectable Effect	20%

In Table 4, it should be noted that the average number of households per cluster (canton) was below the lowest expected size in the initial calculations, with 19.3 households per canton. Also, the intra-cluster

[9] Intuitively, the MDE of a study is the smallest true effect that can be detected with acceptable certainty

[10] The treatment and control groups' compositions define the test being performed.

correlations are generally higher; consequently, having a smaller number of households per cluster will have a smaller effect on the power of the study since a higher intra-cluster correlation implies that it is more efficient to have more clusters and reduce their size for a given sample size. Table 5 shows the corresponding figures using a 10% confidence level.

TABLE 4 SAMPLE SIZE REQUIREMENTS: 5% CONFIDENCE

Test Number	Treatment Group	Control Group	Intra-cluster correlation	Des. Effect	Clusters per condition	Total sample size
80% Power and 5% Confidence						
Average Households per cluster = 19.3						
1	T2	T1	0.05	1.90	23	873
2	T2	T3	0.07	2.25	27	1,036
3	T3	T4	0.08	2.49	30	1,144
4	T5	T4	0.05	1.89	23	869
5	T6	T7	0.04	1.68	20	773

TABLE 5 SAMPLE SIZE REQUIREMENTS: 10% CONFIDENCE

Test Number	Treatment Group	Control Group	Intra-cluster correlation	Des. Effect	Clusters per condition	Total sample size
80% Power and 10% Confidence						
Average Households per cluster = 19.3						
1	T2	T1	0.05	1.90	18	688
2	T2	T3	0.07	2.25	21	816
3	T3	T4	0.08	2.49	23	901
4	T5	T4	0.05	1.89	18	685
5	T6	T7	0.04	1.68	16	609

Given these calculations, we can obtain the number of clusters per highway segment that are needed by taking the maximum number of clusters per segment in each test. For example, segment 2 should have at least 27 clusters since this is the maximum between test 1 and test 2 in Table 4. The results for the required sample per highway segment are presented in

Table 6 in columns 2 and 3. We contrast these results with those in the effective sample at baseline. These results are comparable to those obtained with the EHPM in that if baseline data was available at the sample design stage, these would have been the sample size recommendations. We note that there were considerable deviations in the number of clusters. This is due to the different sizes of the cantons and the availability of enough cantons per segment in the sample frame. For the 10% confidence level, the deviations are smaller, as can be seen in Table 7.

Table 6 Recommended vs. Actual sample size: 80% Power, 5% Confidence

	Design		Actual	
Segment	Clusters per Segment	Sample per Segment	Clusters per Segment	Sample per Segment
T1	23	436	8	243
T2	27	518	13	367
T3	30	572	27	392
T4	30	572	33	610
T5	23	435	23	641
T6	20	386	46	611
T7	20	386	29	586
Total Sample	172	3,306	179	3,450

Table 7 Recommended vs. Actual sample size: 80% Power, 10% Confidence

	Design		Actual	
Segment	Clusters per Segment	Sample per Segment	Clusters per Segment	Sample per Segment
T1	18	344	8	243
T2	21	408	13	367
T3	23	451	27	392
T4	23	451	33	610
T5	18	342	23	641
T6	16	304	46	611
T7	16	304	29	586
Total Sample	135	2,604	179	3,450

To gauge the effect of these deviations on the power of the proposed tests, we calculate the minimum detectable effects for the realized sample using a post-treatment comparison methodology. The equation for the MDE takes the form (from equation 1):

$$\hat{\Delta} = \sqrt{\frac{2\left(1 + (m-1)\widehat{ICC}\right)\left(t_{\alpha/2} + t_{\beta}\right)^2}{mg}\hat{\sigma}^2{}_y}$$

The results are presented in

Table 8. In general, the minimum detectable effect (MDE) for expenditures is above the planned 20% in tests 1, 2, and 3, with test 1 having the largest, meaning that it will be more difficult to detect a 20%

increase in household expenditure if such an increase occurred. At the 10% confidence level (Table 9), the MDE is smaller but still above 20% for tests 1 and 2.

TABLE 8 MINIMUM DETECTABLE DIFFERENCES: 80% POWER, 5% CONFIDENCE

				Actual		
Test	Treatment	Control	Clusters in sample	Sample	Minimum Detectable Effect ($)	Minimum Detectable Effect (%)
1	T2	T1	21	610	$ 60.85	23.9%
2	T2	T3	40	759	$ 59.42	23.4%
3	T3	T4	60	1,002	$ 54.35	21.4%
4	T5	T4	57	1,251	$ 42.41	16.7%
5	T6	T7	76	1,197	$ 40.87	16.1%

TABLE 9 MINIMUM DETECTABLE DIFFERENCES: 80% POWER, 10% CONFIDENCE

				Actual		
Test	Treatment	Control	Clusters in sample	Sample	Minimum Detectable Effect ($)	Minimum Detectable Effect (%)
1	T2	T1	21	610	$ 54.01	21.2%
2	T2	T3	40	759	$ 52.74	20.7%
3	T3	T4	60	1,002	$ 48.24	19.0%
4	T5	T4	57	1,251	$ 37.64	14.8%
5	T6	T7	76	1,197	$ 36.27	14.3%

Given these large MDE's for expenditure, we follow with a power analysis for employment indicators and some intermediate outcomes. The results for employment indicators are presented in Table 10 for number of jobs in which household members participate, number weeks and months worked per year per worker for independent/self-employment and dependent/salaried; Table 11 presents these variables for the total across these categories. These results show that we are able to detect small differences across these variables, so that there is an increase in economic activity around the NTH we are very likely to detect it.

TABLE 10 MDE FOR EMPLOYMENT OUTCOMES: INDEPENDENT AND DEPENDENT WORK

			Mean Outcome at Baseline	Var Outcome at Baseline	Households per Cluster	Total Sample	Intracluster Correlation	Minimum Detectable Difference	Minimum Detectable Difference (%)
Independent Workers									
1	T2	T1	1.17	0.20	25	610	0.05	0.07	5.8%
2	T2	T3	1.19	0.20	17	759	0.01	0.02	2.1%
3	T3	T4	1.16	0.17	15	1002	0.01	0.02	1.5%
4	T5	T4	1.14	0.16	20	1252	0.04	0.04	3.1%
5	T6	T7	1.18	0.19	14	1197	0.06	0.04	3.5%
6	Total		1.17	0.18	19	3450	0.04	0.02	1.8%
Indep. Weeks/Yr per Worker									
1	T2	T1	33.96	183.19	25	610	0.13	3.51	10.3%
2	T2	T3	33.53	177.71	17	759	0.13	2.53	7.5%
3	T3	T4	32.49	187.62	15	1002	0.12	2.00	6.2%
4	T5	T4	31.83	207.34	20	1252	0.15	2.48	7.8%
5	T6	T7	31.30	203.30	14	1197	0.18	2.25	7.2%
6	Total		32.20	199.26	19	3450	0.16	1.47	4.6%
Indep. Months/Yr per Worker									
1	T2	T1	8.54	8.89	25	610	0.01	0.15	1.8%
2	T2	T3	8.28	8.50	17	759	0.05	0.33	3.9%
3	T3	T4	7.99	8.95	15	1002	0.03	0.22	2.7%
4	T5	T4	8.08	9.34	20	1252	0.01	0.13	1.6%
5	T6	T7	7.76	10.06	14	1197	0.00	0.08	1.0%
6	Total		8.05	9.51	19	3450	0.02	0.11	1.4%
Dependent Workers									
1	T2	T1	1.34	0.55	25	610	0.02	0.08	5.9%
2	T2	T3	1.35	0.54	17	759	0.00	0.00	0.0%
3	T3	T4	1.40	0.58	15	1002	0.00	0.00	0.0%
4	T5	T4	1.39	0.54	20	1252	0.00	0.00	0.0%
5	T6	T7	1.44	0.69	14	1197	0.11	0.10	7.1%
6	Total		1.40	0.59	19	3450	0.05	0.05	3.2%
Dep. Weeks/Yr per Worker									
1	T2	T1	32.37	313.70	25	610	0.10	3.98	12.3%
2	T2	T3	31.39	329.51	17	759	0.22	4.37	13.9%
3	T3	T4	32.47	316.62	15	1002	0.19	3.29	10.1%
4	T5	T4	30.58	323.70	20	1252	0.15	3.07	10.0%
5	T6	T7	28.80	312.90	14	1197	0.10	2.07	7.2%
6	Total		30.36	321.52	19	3450	0.13	1.69	5.6%
Dep. Months/Yr per Worker									
1	T2	T1	8.62	14.56	25	610	0.15	1.04	12.1%
2	T2	T3	8.67	15.06	17	759	0.31	1.12	12.9%
3	T3	T4	8.68	15.31	15	1002	0.21	0.76	8.7%
4	T5	T4	8.06	15.94	20	1252	0.10	0.56	6.9%
5	T6	T7	8.13	14.76	14	1197	0.11	0.49	6.0%
6	Total		8.32	15.20	19	3450	0.14	0.39	4.7%

TABLE 11 MDE FOR EMPLOYMENT OUTCOMES: TOTAL WORK

			Mean Outcome at Baseline	Var Outcome at Baseline	Households per Cluster	Total Sample	Intracluster Correlation	Minimum Detectable Difference	Minimum Detectable Difference (%)
Total Job-Workers									
1	T2	T1	1.52	0.65	25	610	0.03	0.11	7.0%
2	T2	T3	1.56	0.65	17	759	0.03	0.07	4.7%
3	T3	T4	1.59	0.77	15	1002	0.02	0.06	3.6%
4	T5	T4	1.53	0.71	20	1252	0.02	0.05	3.5%
5	T6	T7	1.62	0.82	14	1197	0.10	0.11	6.8%
6	Total		1.57	0.74	19	3450	0.06	0.05	3.4%
Total Weeks/Yr per Worker									
1	T2	T1	21.79	278.01	25	610	0.16	4.69	21.5%
2	T2	T3	20.78	279.28	17	759	0.14	3.28	15.8%
3	T3	T4	21.06	266.45	15	1002	0.13	2.55	12.1%
4	T5	T4	20.14	273.76	20	1252	0.11	2.38	11.8%
5	T6	T7	18.79	266.08	14	1197	0.13	2.25	12.0%
6	Total		20.00	273.06	19	3450	0.12	1.48	7.4%
Total Months/Yr per Worker									
1	T2	T1	8.83	9.08	25	610	0.16	0.86	9.7%
2	T2	T3	8.62	9.06	17	759	0.16	0.63	7.3%
3	T3	T4	8.39	9.74	15	1002	0.17	0.56	6.6%
4	T5	T4	8.27	10.31	20	1252	0.12	0.49	6.0%
5	T6	T7	8.09	10.36	14	1197	0.16	0.48	5.9%
6	Total		8.35	10.06	19	3450	0.15	0.32	3.9%

To conclude the post-test power analysis we present the results for some intermediate indicators in Table 12 for the travel times to markets, groceries and health services; and Table 13 for the number of visits per month to the same places. The differences we can detect for the number of visits to markets, grocery, health units, etc., and the number of minutes of the time variables, are small. However, these questions in the survey have a high rate of non-response, so that these results are likely to overstate the power of the tests.

Table 12 MDE for Intermediate outcomes: Travel Times

			Mean Outcome at Baseline	Variance of Outcome at Baseline	Households per Cluster	Total Sample	Intracluster Correlation	Minimum Detectable Difference (Minutes)	Minimum Detectable Difference (%)
Time to Market									
1	T2	T1	28.82	532.14	25	610	0.50	11.61	40.3%
2	T2	T3	37.74	967.68	17	759	0.73	13.79	36.5%
3	T3	T4	35.98	1075.87	15	1002	0.76	12.19	33.9%
4	T5	T4	34.88	992.91	20	1252	0.44	9.25	26.5%
5	T6	T7	33.27	816.37	14	1197	0.29	5.83	17.5%
6	Total		34.03	888.45	19	3450	0.47	5.37	15.8%
Time to Grocery									
1	T2	T1	5.47	22.45	25	610	0.05	0.74	13.5%
2	T2	T3	6.00	27.44	17	759	0.04	0.56	9.4%
3	T3	T4	6.88	43.18	15	1002	0.10	0.90	13.1%
4	T5	T4	7.81	54.45	20	1252	0.18	1.37	17.5%
5	T6	T7	8.74	60.36	14	1197	0.16	1.18	13.5%
6	Total		7.55	49.58	19	3450	0.17	0.77	10.3%
Time to Hospital									
1	T2	T1	40.85	1202.14	25	610	0.18	10.56	25.8%
2	T2	T3	52.16	1855.93	17	759	0.49	15.66	30.0%
3	T3	T4	53.41	2151.85	15	1002	0.44	13.10	24.5%
4	T5	T4	68.38	2948.78	20	1252	0.35	14.11	20.6%
5	T6	T7	69.77	2836.46	14	1197	0.60	15.64	22.4%
6	Total		61.96	2606.65	19	3450	0.47	9.20	14.9%
Time to Health Unit									
1	T2	T1	25.80	375.20	25	610	0.37	8.44	32.7%
2	T2	T3	25.60	520.20	17	759	0.33	6.83	26.7%
3	T3	T4	24.98	487.12	15	1002	0.36	5.63	22.5%
4	T5	T4	28.41	505.67	20	1252	0.35	5.89	20.7%
5	T6	T7	26.87	474.65	14	1197	0.27	4.30	16.0%
6	Total		26.89	484.64	19	3450	0.32	3.29	12.2%

TABLE 13 MDE FOR INTERMEDIATE OUTCOMES: VISITS

			Mean Outcome at Baseline	Var Outcome at Baseline	Households per Cluster	Total Sample	Intracluster Correlation	Minimum Detectable Difference (Times per Month)	Minimum Detectable Difference (%)
Visits to Market									
1	T2	T1	4.50	15.03	25	610	0.05	0.60	13.3%
2	T2	T3	4.29	14.13	17	759	0.01	0.21	4.8%
3	T3	T4	4.56	19.93	15	1002	0.01	0.19	4.2%
4	T5	T4	4.28	14.05	20	1252	0.04	0.34	7.9%
5	T6	T7	3.69	6.25	14	1197	0.06	0.24	6.4%
6	Total		4.12	11.81	19	3450	0.04	0.17	4.2%
Visits to Grocery									
1	T2	T1	18.12	183.07	25	610	0.13	3.51	19.4%
2	T2	T3	17.55	199.73	17	759	0.13	2.68	15.3%
3	T3	T4	17.22	199.19	15	1002	0.12	2.06	12.0%
4	T5	T4	16.78	187.96	20	1252	0.15	2.36	14.1%
5	T6	T7	13.06	110.80	14	1197	0.18	1.66	12.7%
6	Total		15.68	165.14	19	3450	0.16	1.33	8.5%
Visits to Hospital									
1	T2	T1	1.36	0.97	25	610	0.01	0.05	3.7%
2	T2	T3	1.39	1.01	17	759	0.05	0.11	8.0%
3	T3	T4	1.40	1.12	15	1002	0.03	0.08	5.5%
4	T5	T4	1.37	1.03	20	1252	0.01	0.04	3.0%
5	T6	T7	1.51	1.41	14	1197	0.00	0.03	1.9%
6	Total		1.42	1.16	19	3450	0.02	0.04	2.7%
Visits to Health Unit									
1	T2	T1	1.36	0.63	25	610	0.03	0.10	7.2%
2	T2	T3	1.39	0.56	17	759	0.06	0.10	7.1%
3	T3	T4	1.36	0.68	15	1002	0.04	0.07	5.0%
4	T5	T4	1.30	0.59	20	1252	0.03	0.05	4.2%
5	T6	T7	1.50	0.81	14	1197	0.05	0.08	5.3%
6	Total		1.40	0.69	19	3450	0.05	0.05	3.5%

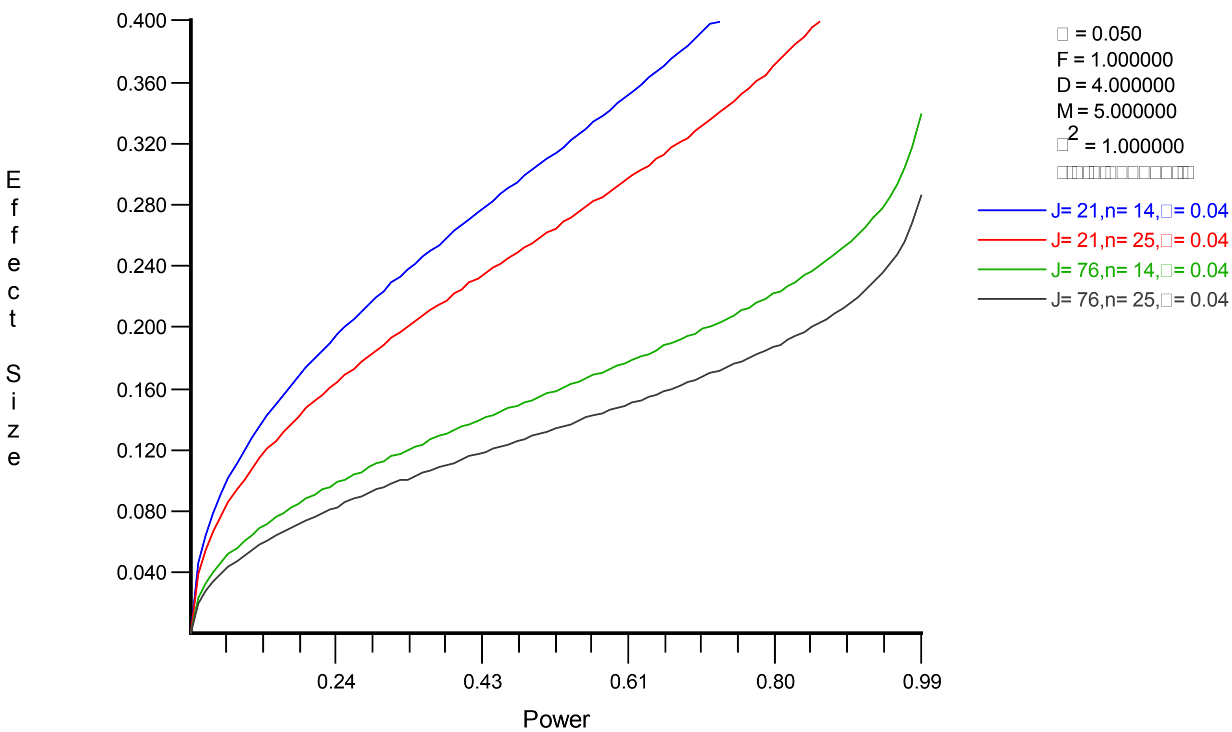
Note that the MDEs above are calculated with baseline data and under the assumption of the post-treatment test. We will see some gains in precision and power by controlling for covariates and the multiple rounds of data available. These two aspects of the design should improve the power of the study and the diminished the effect of attrition in the effective sample size at endline. Next we report the power calculations based on the four rounds (2009-2012) available at the time of writing using the following equation (from equation 6, we call these MDEs with panel adjustment):

$$\hat{\Delta} = \sqrt{\frac{4\left(\hat{\sigma}^2{}_m\left(1 - r_{yy(m)}\right) + m\hat{\sigma}^2{}_g\left(1 - r_{yy(g)}\right)\right)\left(t_{\alpha/2} + t_\beta\right)^2}{mg}}$$

Table 14 shows the power analysis results using all four rounds available and assuming one impact estimate for the full post-treatment period[11]. In general, the situation is similar to that discussed previously, with the tests that involve highway segments 1 and 2 being above the 20% expected effect and the other comparisons safely below the 20 percent MDE that was used as an MDE in the initial sample design. However, this numbers are well above the 6 percent that is now the expected effect. Table 15 disaggregates the MDEs by year post-intervention; the base year is always 2009, and we calculate the MDE in test 1 when comparing 2009 vs. 2010, then 2009 vs. 2011, and so on for each test and each available year. As expected, the MDEs are higher given that we are using less data in these tests vis-à-vis those in Table 14.

Finally, we present Figure 9 with the range of possible effect sizes that can be detected given our sample[12]. For ease of exposition, we present Effect Size as a function of Power for the minimum and maximum cluster size in the tests proposed and the minimum and maximum number of cluster across these tests. The figure shows that for the sample with fewer clusters than expected, the power of the tests for a 20% impact will be lowered.

FIGURE 9 PLAUSIBLE RANGES OF MDES GIVEN CLUSTER SIZE AND NUMBER OF CLUSTERS



---

[11] Note that round 5 of the survey will also increase the number of degrees of freedom available to calculate the impact; this will also reduce the minimum detectable effects.

[12] F is the number of pre-treatment waves, D is the number of post treatment waves, J is the number of cluster, n is the number of observations per cluster and rho is the intracluster correlation. Figure obtained using Optimal Design software Raudenbush and al. (2011)

TABLE 14 MDE WITH PANEL ADJUSTMENT: 5% CONFIDENCE, 3 ROUNDS POST TREATMENT

Test	Treatment	Control	Mean expenditure at Baseline ($)	Households per Cluster	Clusters per condition	Within cluster Correlation across t: $r_{yy(g)}$	Within household Correlation across t: $r_{yy(m)}$	Variance of Impact estimate	Degrees of Freedom	Minimum Detectable Difference ($)	Minimum Detectable Difference (%)
1	T2	T1	265.18	25	11	0.32	0.05	582.05	57	68.77	25.9%
2	T2	T3	264.89	17	20	0.34	0.05	422.62	114	58.09	21.9%
3	T3	T4	253.50	15	30	0.63	0.06	251.25	174	44.66	17.6%
4	T5	T4	246.05	20	29	0.54	0.05	188.50	165	38.69	15.7%
5	T6	T7	255.18	14	38	0.44	0.06	197.24	222	39.52	15.5%

TABLE 15 MDE WITH PANEL ADJUSTMENT: 5% CONFIDENCE, BASELINE VERSUS EACH YEAR

			Year	Households per Cluster	Clusters per condition	Within cluster Correlation across t: $r_{yy(g)}$	Within household Correlation across t: $r_{yy(m)}$	Variance of Impact estimate	Degrees of Freedom	Minimum Detectable Difference ($)	Minimum Detectable Difference (%)
Test 1	Treatment	Control	2010	26.4	11	0.21	0.06	773.69	19	82.17	31.0%
	T2	T1	2011	26.2	11	0.32	0.04	623.57	19	73.76	27.8%
			2012	26.4	11	0.35	0.04	619.27	19	73.51	27.7%
Test 2	Treatment	Control	2010	17.2	20	0.22	0.05	529.11	38	66.15	25.0%
	T2	T3	2011	17.1	20	0.43	0.05	442.75	38	60.51	22.8%
			2012	17.2	20	0.34	0.05	456.97	38	61.47	23.2%
Test 3	Treatment	Control	2010	15.3	30	0.54	0.05	267.30	58	46.59	18.4%
	T3	T4	2011	15.2	30	0.64	0.05	286.88	58	48.26	19.0%
			2012	15.3	30	0.51	0.06	278.55	58	47.56	18.8%
Test 4	Treatment	Control	2010	20.1	29	0.51	0.04	202.89	55	40.63	16.5%
	T5	T4	2011	20.1	29	0.45	0.04	225.93	55	42.87	17.4%
			2012	20.6	29	0.45	0.05	214.92	55	41.81	17.0%
Test 5	Treatment	Control	2010	14.1	38	0.43	0.04	231.05	74	43.15	16.9%
	T6	T7	2011	14.3	38	0.42	0.06	235.01	74	43.52	17.1%
			2012	14.7	38	0.32	0.05	241.60	74	44.13	17.3%

TABLE 16 MDE WITH PANEL ADJUSTMENT: 10% CONFIDENCE, 3 ROUNDS POST TREATMENT

Test	Treatment	Control	Mean expenditure at Baseline ($)	Households per Cluster	Clusters per condition	Within cluster Correlation across t: $r_{yy(g)}$	Within household Correlation across t: $r_{yy(m)}$	Variance of Impact estimate	Degrees of Freedom	Minimum Detectable Difference ($)	Minimum Detectable Difference (%)
1	T2	T1	265.18	25	11	0.32	0.05	582.05	57	60.80	22.9%
2	T2	T3	264.89	17	20	0.34	0.05	422.62	114	51.46	19.4%
3	T3	T4	253.50	15	30	0.63	0.06	251.25	174	39.58	15.6%
4	T5	T4	246.05	20	29	0.54	0.05	188.50	165	34.30	13.9%
5	T6	T7	255.18	14	38	0.44	0.06	197.24	222	35.04	13.7%

TABLE 17 MDE WITH PANEL ADJUSTMENT: 10% CONFIDENCE, BASELINE VERSUS EACH YEAR

			Year	Households per Cluster	Clusters per condition	Within cluster Correlation across t: $r_{yy(g)}$	Within household Correlation across t: $r_{yy(m)}$	Variance of Impact estimate	Degrees of Freedom	Minimum Detectable Difference ($)	Minimum Detectable Difference (%)
Test 1	Treatment	Control	2010	26.4	11	0.21	0.06	773.69	19	72.04	27.2%
	T2	T1	2011	26.2	11	0.32	0.04	623.57	19	64.68	24.4%
			2012	26.4	11	0.35	0.04	619.27	19	64.45	24.3%
Test 2	Treatment	Control	2010	17.2	20	0.22	0.05	529.11	38	58.36	22.0%
	T2	T3	2011	17.1	20	0.43	0.05	442.75	38	53.39	20.2%
			2012	17.2	20	0.34	0.05	456.97	38	54.24	20.5%
Test 3	Treatment	Control	2010	15.3	30	0.54	0.05	267.30	58	41.19	16.2%
	T3	T4	2011	15.2	30	0.64	0.05	286.88	58	42.67	16.8%
			2012	15.3	30	0.51	0.06	278.55	58	42.05	16.6%
Test 4	Treatment	Control	2010	20.1	29	0.51	0.04	202.89	55	35.91	14.6%
	T5	T4	2011	20.1	29	0.45	0.04	225.93	55	37.90	15.4%
			2012	20.6	29	0.45	0.05	214.92	55	36.96	15.0%
Test 5	Treatment	Control	2010	14.1	38	0.43	0.04	231.05	74	38.19	15.0%
	T6	T7	2011	14.3	38	0.42	0.06	235.01	74	38.51	15.1%
			2012	14.7	38	0.32	0.05	241.60	74	39.05	15.3%

b. Sample Conditions for Continuous Treatment

In this case, the treatment is the change in time it takes to get to the nearest market[13] resulting from the improvement of the highway. Households from segment *j* will benefit from improvements in segment *i* (which may or may not be equal to *j*) if they use that part of the highway to send their products to the market. Therefore, the treatment can be interpreted as the change in time to the market.

To calculate the sample size necessary to detect a given effect, we can estimate an equation and construct a test for the null hypotheses that the effect is zero (two-tailed) and use this expression to solve for the sample size or, conversely, for the minimum detectable effects given our sample. In addition, this test needs to take into account the intra-cluster correlation so as to not inflate the power of the tests. In its simplest form, we test the coefficient of the treatment in a linear regression of expenditure on the treatment and a constant only using the baseline data.

The equation of interest is:

$$Y_i = \alpha + \theta T_i + \varepsilon_i$$

Murray (1980) shows that with many degrees of freedom, $t = 2.80$ guarantees a power of 80%. Hence, we need to look for the conditions that would guarantee obtaining a t value of at least that magnitude.

$$\hat{t} = \frac{|\hat{\theta}|}{\hat{\sigma}_\theta} \geq 2.80$$

Hence, it is required that

$$|\hat{\theta}| \geq 2.80\sigma_{\hat{\theta}}$$

More generally, for a power of $1 - \beta$ and a confidence at the α level,

$$|\hat{\theta}| \geq (t_{1-\frac{\alpha}{2}} + t_{1-\beta})\sigma_{\hat{\theta}}$$

Note that the OLS estimator of the coefficient of T is

$$\theta = \frac{cov(Y,T)}{var(T)} = \frac{\sigma_{YT}}{\sigma_T^2}$$

and that the variance of the outcome can be expressed as

$$var(Y) = \theta^2 var(T) + var(\varepsilon) \rightarrow \sigma_y^2 = \theta^2\sigma_T^2 + \sigma_\varepsilon^2$$

---

[13] Defined as any town/city with a population of 25,000 inhabitants or more

Substituting the OLS estimator in the equation above,

$$\sigma_Y^2 = \left(\frac{\sigma_{YT}}{\sigma_T}\right)^2 + \sigma_\varepsilon^2$$

$$1 = \left(\frac{\sigma_{YT}}{\sigma_T \sigma_Y}\right)^2 + \frac{\sigma_\varepsilon^2}{\sigma_y^2}$$

And

$$(1 - \rho_{YT}^2)\sigma_Y^2 = \sigma_\varepsilon^2$$

Where $\rho_{YT} = \frac{\sigma_{YT}}{\sigma_T \sigma_Y}$ is the correlation coefficient between Y and T

Hence, the variance of the OLS estimate of θ can be estimated by:

$$var(\hat{\theta}) = \frac{\hat{\sigma}_\varepsilon^2}{\Sigma_i (T_i - \bar{T})^2} = \frac{(1 - \hat{\rho}_{YT}^2)\hat{\sigma}_Y^2}{n\hat{\sigma}_T^2}$$

To construct the test, start by asserting:

$$t_{1-\frac{\alpha}{2}} + t_{1-\beta} = \frac{|\hat{\theta}|}{\sqrt{\frac{(1 - \hat{\rho}_{YT}^2)\hat{\sigma}_Y^2}{n\hat{\sigma}_T^2}}}$$

Taking into account the intra-cluster correlation:

$$t_{1-\frac{\alpha}{2}} + t_{1-\beta} = \frac{|\hat{\theta}|}{\sqrt{\frac{(1 - \hat{\rho}_{YT}^2)(1 + (1-m)\widehat{ICC})\hat{\sigma}_Y^2}{mg\hat{\sigma}_T^2}}}$$

where m is the number of households per cluster and g is the number of clusters in the sample, so that[14] $n = mg$.

The sample size requirements are given by:

$$g = \frac{(1 + (1-m)\widehat{ICC})\left(t_{1-\frac{\alpha}{2}} + t_{1-\beta}\right)^2}{m\hat{\theta}^2} \frac{(1 - \hat{\rho}_{YT}^2)\hat{\sigma}_Y^2}{\hat{\sigma}_T^2}$$

---

[14] Note that in this case, $n \neq 2mg$ since we have do not have treatment and control conditions, and g is the total number of clusters in the sample.

The estimates $\hat{\theta}, \hat{\sigma}_Y^2, \hat{\sigma}_T^2, \hat{\rho}_{YT}^{\,2}$, and $\widehat{ICC}$ for each segment can be obtained from the baseline survey to compute the sample size required.

In addition, the minimum detectable effect is given by:

$$|\hat{\theta}| = \sqrt{\frac{(1-\hat{\rho}_{YT}^2)(1+(1-m)\widehat{ICC})\left(t_{1-\frac{\alpha}{2}}+t_{1-\beta}\right)^2}{mg\hat{\sigma}_T^2}\hat{\sigma}_Y^2}$$

We can determine the lowest $\hat{\theta}$ that we can detect given our realized sample in each segment of the highway. Table 18 shows the results of these calculations. As was the case with the discrete approach, the MDEs in the first three highway segments are high. However, the advantage of the continuous treatment approach is that it uses the complete sample across all segments of the highway. Given this sample, we can detect an effect of a $23.36 per month in expenditures for each hour decrease in the traveling times to markets due to the highway. This corresponds to 9.2 percent of the observed average monthly household expenditure at baseline.

As we did in the previous sections, we present the power analysis for some intermediate outcomes and employment outcomes. For the sake of brevity we present the results for visits to markets, visit to health units, and number of jobs in which household members participate, total number weeks and months worked per year per worker. The results show that the sample can detect small size relations between these variables.

The sample size required for a continuous treatment approach is in general smaller than that needed for the discrete treatment case. These calculations show that the discrete sample size also complies with the sample requirements in the continuous case. One must also keep in mind that the test in the continuous case focuses on only one parameter; if a more complex functional form with multiple rounds and covariates were used, the chances of detecting an effect will be larger or, conversely, the minimum detectable effects could be smaller and the estimated impacts more precise.

Even though some of the parameters obtained from EHPM and the census in the initial design were different than those obtained from the baseline survey, the realized sample is well powered to detect the expected changes in expenditure due to the highway construction. The tests that involve segments 1 and 2 of the highway present the largest minimum detectable effects, so that if any effects were detected below the minimum, these would be at the expense of lower power.

It must be kept in mind that changes in the timing of the construction of the highway's sections have implications on the size of the effects that can be detected given the timing of the surveys. In the final analysis, it will be important to take these differences into account and explore the possibility of exploiting a finer variation of this timing (if reliable data on the timing of each part of the highway were available) to increase the precision and power of these estimates.

c. Conclusions of Power Analysis

In summary, the results across methodologies show that the survey is powered to detect only large effects using the discrete treatment methodology (RD), from 15 to 29 percent changes in income/consumption depending on the segments being compared, which is above of the 6 percent income increase revised target expected by MCC.

Using the continuous treatment methodology it is powered to detect an effect equal to 9.5 percent of baseline income per hour of time gained. We note that even though this seems a large gain, the impact estimate using this methodology is a combination of the differences in means and the detected "structural" relationship between the outcome and the travel time variables. This implies that we will be able to detect small impacts using this sample using the continuous treatment. For example, we could detect an impact as low as 6% of the baseline expenditure if the average reduction in travel times is 38 minutes (which is considerably large), depending on the different gains household might have (since these gains are in the order of minutes). The ex-post precision of the estimate will depend on the correlation of the between travel time across time and the variance of the parameter estimate that relates the outcome variables with travel time. Without making heroic assumption about these variables one cannot gauge exactly the power of the tests ex-ante. One of the advantages of this research is that it could serve as a benchmark which future studies could use to conduct ex-ante power analysis.

TABLE 18 MDE FOR CONTINUOUS TREATMENT APPROACH: 80% POWER, 5% CONFIDENCE

Highway Segment	$\widehat{\sigma}_T^2$	$\widehat{\sigma}_y^2$	$\widehat{\rho}_{YT}$	ICC	$\widehat{\theta}$	Sample size (n)	Households per Cluster (m)	Clusters (g)	Degrees of Freedom	Design Effect	Minimum Detectable $\widehat{\theta}$ ($ per Hour)	Minimum Detectable $\widehat{\theta}$ (% Baseline Expenditure per Hour)
T1	0.35	37,889	-0.01	0.110	-3.4	236	29.2	8	7	4.111	143.44	55.2%
T2	0.22	42,106	-0.01	0.036	-5.2	364	29.0	13	12	1.994	97.26	36.2%
T3	0.13	46,279	-0.22	0.104	-130.8	378	14.9	27	26	2.443	130.80	50.1%
T4	0.30	35,918	-0.10	0.074	-36.1	596	18.4	33	32	2.281	61.21	24.6%
T5	0.27	32,515	-0.08	0.027	-28.7	624	27.8	23	22	1.721	52.66	21.6%
T6	0.28	36,703	-0.11	0.074	-40.2	602	13.2	46.0	45.0	1.904	57.72	23.7%
T7	0.41	38,785	-0.06	0.002	-19.3	577	20.1	29.0	28.0	1.034	37.46	14.0%
Total	0.30	37,917	-0.08	0.050	-28.3	3,377	19.3	179.0	178.0	1.917	23.36	9.2%

TABLE 19 MDE FOR CONTINUOUS TREATMENT APPROACH: INTERMEDIATE OUTCOMES AND EMPLOYMENT

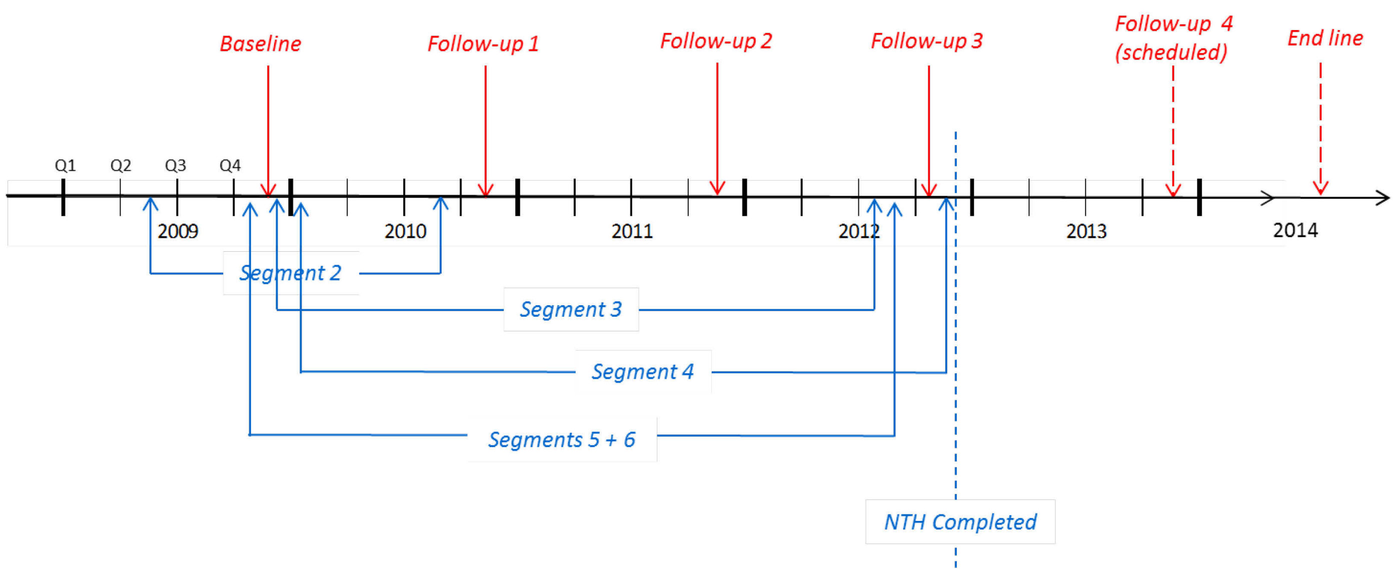
Highway Segment	$\hat{\sigma}_T^2$	$\hat{\sigma}_y^2$	$\hat{\rho}_{YT}$	ICC	$\hat{\theta}$	Sample size (n)	Households per Cluster (m)	Clusters (g)	Degrees of Freedom	Design Effect	Minimum Detectable $\hat{\theta}$	Minimum Detectable $\hat{\theta}$ (% Baseline)
Visits to Markets												
T1	0.35	19	-0.15	0.011	-1.1	236	29.2	8	7	1.297	1.77	0.7%
T2	0.20	13	-0.13	0.071	-1.1	364	29.0	13	12	2.998	2.17	0.8%
T3	0.11	16	-0.06	0.000	-1.0	378	14.9	27	26	1.000	1.71	0.7%
T4	0.31	22	-0.14	0.031	-1.3	596	18.4	33	32	1.536	1.21	0.5%
T5	0.27	4	-0.08	0.000	-0.3	624	27.8	23	22	1.000	0.43	0.2%
T6	0.28	11	0.01	0.000	0.1	602	13.2	46.0	45.0	1.000	0.72	0.3%
T7	0.42	2	-0.06	0.110	-0.2	577	20.1	29.0	28.0	3.103	0.46	0.2%
Total	0.31	12	-0.12	0.036	-0.7	3,377	19.3	179.0	178.0	1.666	0.38	0.1%
Visits to Health Unit												
T1	0.35	1.0	-0.02	0.000	0.0	236	29.2	8	7	1.000	0.36	0.1%
T2	0.20	0.4	-0.07	0.000	-0.1	364	29.0	13	12	1.000	0.24	0.1%
T3	0.11	0.6	-0.06	0.091	-0.1	378	14.9	27	26	2.266	0.51	0.2%
T4	0.31	0.7	0.06	0.005	0.1	596	18.4	33	32	1.082	0.18	0.1%
T5	0.27	0.5	-0.01	0.047	0.0	624	27.8	23	22	2.248	0.23	0.1%
T6	0.28	0.8	-0.08	0.058	-0.1	602	13.2	46.0	45.0	1.713	0.26	0.1%
T7	0.42	0.8	0.00	0.050	0.0	577	20.1	29.0	28.0	1.945	0.23	0.1%
Total	0.31	0.7	0.00	0.051	0.0	3,377	19.3	179.0	178.0	1.926	0.10	0.0%
Total Job-Workers												
T1	0.35	0.7	0.03	0.048	0.0	236	29.2	8	7	2.338	0.47	0.2%
T2	0.20	0.6	-0.13	0.032	-0.2	364	29.0	13	12	1.907	0.38	0.1%
T3	0.11	0.7	0.01	0.025	0.0	378	14.9	27	26	1.345	0.41	0.2%
T4	0.31	0.8	0.03	0.028	0.0	596	18.4	33	32	1.486	0.24	0.1%
T5	0.27	0.6	-0.04	0.014	-0.1	624	27.8	23	22	1.377	0.20	0.1%
T6	0.28	1.0	0.07	0.119	0.1	602	13.2	46.0	45.0	2.456	0.33	0.1%
T7	0.42	0.6	0.00	0.035	0.0	577	20.1	29.0	28.0	1.664	0.19	0.1%
Total	0.31	0.7	-0.02	0.058	0.0	3,377	19.3	179.0	178.0	2.058	0.11	0.0%

Total Weeks/Yr per Worker												
T1	0.35	276	-0.14	0.219	-4.5	236	29.2	8	7	7.170	16.02	6.2%
T2	0.20	281	-0.23	0.163	-9.8	364	29.0	13	12	5.558	13.58	5.1%
T3	0.11	279	-0.13	0.158	-7.1	378	14.9	27	26	3.192	12.75	4.9%
T4	0.31	258	-0.23	0.150	-7.8	596	18.4	33	32	3.617	6.29	2.5%
T5	0.27	289	-0.14	0.062	-4.8	624	27.8	23	22	2.651	6.10	2.5%
T6	0.28	262	-0.23	0.108	-9.0	602	13.2	46.0	45.0	2.321	5.28	2.2%
T7	0.42	274	-0.12	0.180	-2.9	577	20.1	29.0	28.0	4.424	6.42	2.4%
Total	0.31	273	-0.17	0.117	-5.4	3,377	19.3	179.0	178.0	3.130	2.47	1.0%
Total Months/Yr per Worker												
T1	0.35	9.4	-0.30	0.190	-1.6	236	29.2	8	7	6.348	2.68	1.0%
T2	0.20	8.7	-0.09	0.151	-0.6	364	29.0	13	12	5.222	2.38	0.9%
T3	0.11	9.4	-0.24	0.193	-2.4	378	14.9	27	26	3.678	2.45	0.9%
T4	0.31	10.0	-0.15	0.168	-0.9	596	18.4	33	32	3.919	1.31	0.5%
T5	0.27	10.7	-0.13	0.088	-0.8	624	27.8	23	22	3.367	1.32	0.5%
T6	0.28	10.1	-0.07	0.093	-0.4	602	13.2	46.0	45.0	2.131	1.02	0.4%
T7	0.42	9.9	-0.13	0.171	-0.6	577	20.1	29.0	28.0	4.254	1.20	0.4%
Total	0.31	10.1	-0.16	0.148	-0.9	3,377	19.3	179.0	178.0	3.712	0.52	0.2%

4.6. Timeframe

The data collection activities were performed during the November to February (of the following year) periods, starting in 2009. The proposed design uses 6 survey waves, with the last one measuring longer terms effects expected to take place in November 2014.

FIGURE 10 TIMELINE OF ACTIVITIES



4.6.1. Justification for Proposed Exposure Period to Treatment

In September, 2012, all the segments of the NTH were constructed. However, by the time of each follow-up, each segment had different levels of advancement; thus, households benefiting from the improved roads experienced limited increases in market access since the improved road might not have been finished near the market that the household members needed to travel to. By the time of the proposed endline survey in 2014, all beneficiary households will have access to the improved roads for different periods of time (temporal variation) and different accessibility (spatial variability).

Given this varied exposure time, it is important to have multiple data points to estimate the long-term and short-term impacts of the road. For example, farmers may need more experience marketing their crops and developing business relations to realize the full potential of enhanced access to markets. In addition, the variation of a household's accessibility resulting from the construction process is used in the continuous approach discussed in previous sections. In estimating the impact of the NTH, we need to relate these continuous changes in accessibility over time to changes in relevant outcomes. Timely measures of outcomes require more data collection points.

Sample attrition makes the need for multiple survey rounds evident: if the sample size drops over time, so does its power to detect any impact resulting from the intervention. The number of survey rounds can help us mitigate the adverse effects of attrition, which was considerably high in the first follow-up survey. Longer exposure periods imply that households are less likely to drop from the sample because i) in rounds closer to the initial round, they tend to be easier to locate and ii) we allow more time for return migration, effectively reincorporating households in the later periods. Finally, we gain variability over time, which partially compensates for the loss of cross-sectional variation due to attrition, making the impact estimates more precise.

These are the same reasons why we propose to have the endline in 2014, instead of waiting more time. Allowing more time for the endline would make it more difficult to find the households and would make it more likely that other interventions in are put in place and distort the impact evaluation.

In summary, having more data points per household will increase the degrees of freedom, which is important because any analysis with low degrees of freedom runs the risk of finding false negatives (i.e., finding "no effect" when in reality there is an effect). Having more data points and more frequent allows us to maintain the statistical power of the survey while tracking the impact of the NTH at different points in time and maintaining contamination from other interventions and sample attrition in check.

5. Data Sources and Outcome Definitions

The impact evaluation uses household surveys and community surveys. As detailed in the previous section, a random stratified sample from the census was drawn from the cantons in the NTH's zone of influence.

The household survey for the connectivity impact evaluation interviewed 3,450 households at baseline. As Table 20 shows, there was high attrition in the first two follow-up surveys (2010 and 2011), with few of the households lost in the follow-up being regained in 2011. In the 2012 round, we implemented a farther-reaching tracking of the baseline households and were able to recuperate more households, bringing the effective sample size in the 2012 survey to 3,065 households. For the endline survey in 2014, we will continue to track down the baseline households that are still missing and, to the extent possible, use methods that are robust to unbalanced panel data in the final analysis.

TABLE 20 EFFECTIVE HOUSEHOLD SAMPLE BY YEAR

Survey Year	Effective Sample	Total Attrition	Rolling Attrition
2009	3,450	-	-
2010	2,904	16%	16%
2011	2,935	15%	-1%
2012	3,065	11%	-4%

Table 21 shows the main impacts and outcome indicators in the monitoring and evaluation plan for the connectivity project [(MCC, 2012)]. We note that the traffic survey is still pending and so that the traffic variables are conditional on the implementation of that survey. Other indicators that will be used in the evaluation where discussed in section 4.2.

TABLE 21 MAIN IMPACT AND OUTCOME INDICATORS IN M&E PLAN

Level	Indicator	Definition	Target	Data Source
Impact	Increase in income of households near the Northern Transnational Highway	% Increase in income of households within 2km of the Northern Transnational Highway	6%	Household Survey
Impact	Land prices along the Northern Transnational Highway	Average price of land 2km on either side of the Northern Transnational Highway (weighted average of all road sections to be opened or improved), per m2	from 3.22 $\$/m^2$ to 3.40 $\$/m^2$	Household Survey
Outcome	Average annual daily traffic	The average number and type of vehicles per day, averaged over different times (day and night) and over different seasons to arrive at an annualized daily average.	from 270 to 962 vehicles per day	Traffic Survey

5.1.New and Existing

5.1.1. Quantitative – Household Survey

The **baseline** and **endline** survey questionnaire includes two sections – one (including questions about household income and agricultural productivity) that is answered by the male head of household who is interviewed by a male survey taker and one (including questions about household demographics, time allocation, and expenses) which is answered by a female in the household, i.e. spouse or female household head, who is interviewed by a female survey taker. The survey has detailed sections for each of the outcomes to be evaluated, both intermediate and final indicators. In addition, to be able to control for accessibility to markets, each of the surveyed households was geo-referenced. If the appropriate persons are not present at the time of the first visit, enumerators make an appointment and return again to interview the appropriate person, provided that this return visit is possible within the time that the survey team will be in the area. When possible, a second adult can also be included in the interview process, particularly for the questions related to work and agricultural output. The survey is designed to take between 1 and 1 ½ hours for each questionnaire (i.e. male and female).We will administer the sex differentiated surveys at baseline and endline to be able, to recuperate some of the information that was excluded in the condensed follow-up questionnaires; such as, participation in social programs, agricultural assets, animal rearing, etc. This strategy permits the informant that is the most knowledgeable of specific aspects of the household activities to provide more detailed information without increasing the time burden of the survey.

For the follow-up questionnaires, a condensed version of the previous questionnaire was administered. In this version, the main themes of the survey are presented and administered once per household. The informants are allowed to change during the interview so that the person most informed person about each topic can answer that particular section.

5.1.1. Quantitative and Qualitative - Community Survey

The community survey was administered to key informants in communities where selected households live; each section of the survey was administered to the better-informed informant. For example, questions related to the community infrastructure and prices were administered to the major or community leader, questions related to the health of community members were administered to the health center director, and questions related to education were administered to the school director. The community survey gathered information about the local economy (including price levels for food, basic commodities, wages, etc.), community infrastructure, and access to key markets and social services. The goal of the surveys is to provide some context for the information gathered in the household surveys, to track community-level changes that may affect outcomes, and to reduce the required length of the household survey questionnaire.

The qualitative analysis for the impact evaluation will use this information and open ended questions in the endline surveys to provide some context for the impact pathways previously discussed.

6. Analysis Plan

This section spells out the econometric analysis that will be implemented for the RD and Continuous Approaches described in Sections 4.3.1.

6.1.RD Approach

In this approach, we compare households in adjacent segments where the NTH is rehabilitated in different periods (or is not rehabilitated at all). Consider the following treatment assignments (k) between pairs of segments, where one is assigned to a treatment group (S_k) and the other is used as a control (Table 22).

TABLE 22 TREATMENT ASSIGNMENT BY SEGMENTS

Treat Assignment	Treatment Segment (S_k)	Control Segment
k=1	Segment 2	Segment 1
k=2	Segment 2	Segment 3
k=3	Segment 3	Segment 4
k=4	Segment 5	Segment 4
k=5	Segment 6	Segment 7

The effect of the treatment in each treatment assignment can estimated through:

$$Y_{it} = \sum_{t=1}^{T} \beta_t D_t S_k + \sum_{t=1}^{T} \gamma_t D_t + \theta X_{it} + \alpha_i + \varepsilon_{it} \qquad \text{if } i \in W_k, k = 1, \ldots, T.$$

where D_t are time period indicator variables for t=1,..,T; α_i are household-specific fixed effects (that capture any time-invariant unobservable heterogeneity), X_{it} is vector of control variables, and $\varepsilon_{it} \sim D(0, \sigma^2)$ is an error term. W_k denotes the set of households in each k treatment assignment for which we estimate separate regressions. Note that while W_k might include all households in both segments of each treatment assignment, this sample might also be restricted to households in a buffer area around the segment discontinuity.

Due to the fixed-effects specification, these coefficients exploit the within-household variation among treatment and control groups across periods and each comparison group *k*. Therefore, this methodology allows for the possibility of the NTH having differential effects within comparison groups and time periods.

6.2.Continuous Approach

In the continuous treatment, rather than exploiting the discontinuity between segments, the aim is to capture the impact of reductions in travel time on income and other outcomes. For this purpose, we estimate the following regression:

$$Y_{ist} = \sum_{t=1}^{T} \beta_t V_{ist} + \sum_{t=1}^{T} \gamma_t D_t + \theta X_{ist} + \alpha_i + \varepsilon_{ist}$$

where Y_{ist} is the income of the i-th household in segment s and time t, D_t are indicator variables for each time period, X_{ist} is a vector of control variables, α_i are household-specific fixed effects (to account for time-invariant observables and unobservables), and $\varepsilon_{ist} \sim D(0, \sigma^2)$. V_{ist} is travel time to a relevant market for each household estimated using raster analysis. This approach considers overall reductions in travel time as the NTH rehabilitation progresses. However, it also takes into account that households will be differentially affected depending on their specific location (because segments are improved in different periods). In this light, the coefficient of interest is β_t, which measures the impact of reducing transportation time by one unit in period *t*.

With this methodology, we can also derive estimates that can be compared with those calculated through the RD approach. Consider the same treatment assignment in Table 22. We can estimate the following regression:

$$Y_{ist} = \sum_{k=1}^{5} \sum_{t=1}^{T} \beta_{kt} D_k V_{ist} + \sum_{t=1}^{T} \gamma_t D_t + \sum_{k=1}^{T} \delta_k D_k + \theta X_{sit} + \alpha_i + \varepsilon_{ist}$$

This regression includes an interaction with D_k which allows us to estimate a set of coefficients β_{kt} for each comparison group *k*. Then we can use the average travel time for each segment in different periods to estimate the average treatment effect in each comparison group. For example, consider comparison group *k=1*, where we compare outcomes in segments 2 and 1. The differences in income attributable to the NTH in period *t=T* are: $-\beta_{1T}(\bar{V}_{s=2,t=T} - \bar{V}_{s=1,t=T})$.

These calculations are not strictly equivalent because, as mentioned, both methods exploit different sources of variability. However, we provide this alternative specification to provide rough comparisons of the results under both methodologies.

7. Monitoring Plan

7.1. Adherence to Treatment and Control Areas

The multiple survey rounds have allowed us to address deviations from the project implementation. Indeed, more certain information regarding the dates of future road construction would have been beneficial in minimizing the number of intermediate survey rounds and in clarifying the treatment assignment at each point in time. For example, if we were sure that 50% of the project would have been completed in the middle point of the study, we might have decided to have only one intermediate survey collected. However, we were aware of the potential for changes in the timing of different stages of the project (which eventually happened). Thus, we planned on more periodic waves that would allow us to better accommodate the uncertainty in the implementation timeline while still being able to disentangle the trajectory of the intervention over time.

In the end, more detailed information on the construction of the NTH will allow us to exploit the variability in accessibility further by assigning households to finer treatment and control groups depending on the month in which the household was surveyed and the state of the highway construction during that month.

8. Administrative

8.1. Summary of Institutional Review Board requirements

Following IFPRI Institutional Review Board Guidelines, this impact evaluation would bear minimal risk. These guidelines consider minimal risk to occur when "the probability and magnitude of harm or discomfort anticipated in the proposed research are not greater, in and of themselves, than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests". In particular, this impact evaluation only relies on "research employing survey, interview, oral history, focus group methods", which falls into the minimal risk category.

8.2. Data Access, Privacy, and Documentation Plan

We will produce cleaned raw datasets that follows MCC's guidelines for public use data, including programming syntax used to clean the datasets for documentation purposes.

A full set of documentation for each survey will be provided. The raw data and the data used for the final analysis will be provided. A public use version of analysis data files will be provided. The publicly available version will be anonymized, and thus free of personal or geographic identifiers that would permit identification of individual respondents or their household members. In addition, we will exclude variables that introduce reasonable risks of deductive disclosure of the identity of individual subjects.

In order to facilitate access to and usability of data, all datasets delivered to MCC will be accompanied by completed documentation in the form of standardized metadata using the International Household Survey Network's (IHSN) Metadata Editor.

8.3. Dissemination Plan

We will produce an interim report, tentatively scheduled for October 2013 that will include the analysis of the first four rounds of the survey; a follow-up report will be drafted for November 2014 and revised for January 2015, which will include the analysis of first 5 waves of the household survey. In addition, a report detailing the long term impacts of the connectivity project will be produced after the endline survey.

The reports will follow a template agreed upon with MCC. The reports will validate evaluation design and revise the power analysis when necessary to verify the appropriateness of the effective sample for the impact evaluation. Estimates for figures needed in the ERR models will be included. Finally, the results will be compiled in an academic paper to be published in policy and development journals.

Presentation dissemination efforts will include: presentation of the report(s) to MCC Headquarters staff, presentation in MCC workshops, presentation of findings and key recommendations to local stakeholders, and presentation of the findings in other international development conferences.

8.4.Evaluation Team Roles and Responsibilities

Máximo Torero will manage the technical IFPRI team, which will work closely together on the connectivity impact evaluation. He will lead the technical review, be the primary point of contact for the evaluation, travel to the field to oversee data collection, lead his teams in data management, cleaning, analysis, and reporting, and, finally, will present findings. Mr. Eduardo Nakasone and Mr. Miguel Almánzar will work closely with Dr. Torero by leading the data collection, analysis, and dissemination of results. Finally, Ms. Maribel Elias will have limited involvement as a geographer to support result mapping.

Dr. Máximo Torero has 17 years of experience in research and project management. He is an expert in designing and applying innovative randomized experimental strategies that measure the impacts of development programs across sectors, focusing on infrastructure, information communication technology, and market access. Dr. Torero is the Division Director of IFPRI's Markets, Trade and Institutions Division, where he leads and conducts research with a special emphasis on monitoring and impact evaluation of infrastructure and rural development interventions in urban and peri-urban areas. He has applied his research and experimental designs in various MCC countries, including El Salvador, Mozambique, Peru, Tanzania, Uganda, and Zambia. In addition to working as Program Manager on this MCC El Salvador impact evaluation, Dr. Torero's recent work includes: leading a DFID/CARE-funded project to implement a quasi-experimental impact evaluation of access to infrastructure in the Northwestern region of Bangladesh; leading an Inter-American Development Bank (IADB) funded quasi-natural experiment of access to telephone services and household income in poor rural areas of Peru; and a World Bank-funded regression discontinuity design impact evaluation of rural electrification in Peru. Dr. Torero received his MA and PhD in Economics from UCLA. Dr. Torero is a native Spanish speaker and is fluent in English.

Mr. Eduardo Nakasone is currently a Senior Research Analyst at IFPRI, where his research focuses on the impact of infrastructure, particularly Information and Communication Technologies, on rural development. He has more than 10 years of experience in economic development research, including research on the impact of CAFTA on poverty distribution and growth in El Salvador. For his dissertation, Mr. Nakasone is currently analyzing the impact of access to market price information on marketing outcomes among small farmers in the central highlands of Peru through a randomized controlled trial. He has also evaluated the effect of the privatization of electricity distribution companies in rural Peru and the effectiveness of an entrepreneurship training program among women in urban areas of Peru. In El Salvador, since 2010 he has actively participated on the IFPRI team for connectivity and electrification impact evaluations of the MCC-FOMILENIO projects. Mr. Nakasone is a PhD candidate in the Agricultural and Resource Economics Department of the University of Maryland-College Park and received his MS degree from this institution. He is a native Spanish speaker, is fluent in English, and is a conversational Japanese speaker.

Mr. Miguel Almánzar is currently a Senior Research Analyst at IFPRI, where his research focuses on the effects of public infrastructure and services on rural development. For his dissertation, Mr. Almánzar is currently analyzing the distributional impact of infrastructure provision in El Salvador. He is also evaluating the impact of nutritional and agricultural extension services in Honduras. In El Salvador, since 2012 he has actively participated on the IFPRI team for connectivity and electrification impact evaluations of the MCC-FOMILENIO projects and has been a part of the team working on the impact evaluation of the water and sanitation intervention in Northern El Salvador since 2011. Mr. Almánzar is a PhD candidate in the Agricultural Economics and Rural Development Department of the University of Göttingen and received his MA degree in economics from the University of Maryland-College Park. He is a native Spanish speaker and is fluent in English and French.

Ms. Maribel Elías has been a Research Analyst and Geographic Information Systems (GIS) specialist in IFPRI's Markets Trade and Institutions Division since 2007. Her current research focuses on applying Spatial Analysis and GIS to understand production, marketing, and location decisions among small farmers in developing countries. Her main areas of interest lie in understanding the spatial dynamics of regions and populations, improving measures of accessibility to markets, and enhancing spatial targeting of government programs to improve agricultural productivity. Ms. Elías earned her MA in Geography at San Diego State University, where her thesis analyzed patterns of social indicators, such as education, in regions in Peru over time. She is a native Spanish speaker and is fluent in English.

8.5.Budget

The costs of the connectivity activity ascend to 269 million, as mentioned before. The impact evaluation activities amount to $549,000 in evaluators cost and $1.03 million in data collection activities; with the total impact evaluation cost representing a 0.59 percent of the total cost of improving the NTH[15].

[15] These figures do not include the endline (expected in 2014) survey and related costs.

9. References

Dirección General de Estadística y Censos. (2011, 2012). *Encuesta de Hogares de Propósitos Múltiples.* Retrieved July 2013, from http://www.digestyc.gob.sv/index.php/temas/des/ehpm.html

Ahmed, R., & Hossain, M. (1990). *Development impact of rural infrastructure. .* Washington, D.C.: International Food Policy Research Institute Research Report No. 83.

Antle, J. M. (1983). Infrastructure and aggregate agricultural productivity: International evidence. *Economic Development and Cultural Change, 31*(3), 609–619.

Antle, J. M. (1984). Human capital, infrastructure, and the productivity of Indian rice farmers. *Journal of Development Economics, 14*(1/2), 163–181. .

Aschauer, D. A. (1989). Is public expenditure productive? . *Journal of Monetary Economics , 23*(2), 177-200.

Aschauer, D. A. (1989). Public Investment and productivity growth in the group of seven. . *Economic Perspectives , 13* (5), 17–25.

Binswanger, H. P., Khandker, S. R., & Rosenzweig, M. P. (1993). How infrastructure and financial institutions affect agricultural output and investment in India. *Journal of Development Economics, 41*(2), 337–366.

Dercon, S., Gilligan, D. O., Hoddinott, J., & Woldehanna, T. (2009). The impact of agricultural extension and roads on poverty and consumption growth in fifteen Ethiopian villages. *American Journal of Agricultural Economics, , 91*(4), 1007-1021.

Dong, X. (2000). Public investment, social services and productivity of Chinese household farms. *Journal of Development Studies, 36*(3), 100–122.

Fan, S., Hazell, P., & Thorat, S. (1999). *Linkages Between Government Spending, Growth, and Poverty in Rural India.* Washington, D.C.: International Food Policy Research Institute Research Report 110.

Garcia-Mila, T., McGuire, T. J., & Porter, R. H. (1996). The effect of public capital in state-level production functions reconsidered. *Review of Economics and Statistics, 78* (1), 177–180.

Gibson, J., & Rozelle, S. (2003). Poverty and access to roads in Papua New Guinea. *Economic Development and Cultural Change , 52*(1), 159–185.

Hirschmann, A. O. (1959). *The strategy of economic development.* (2nd ed.). New Haven, Conn.: Yale University Press.

Holtz-Eaakin, D. (1994). Public-sector capital and the productivity puzzle . *Review of Economics and Statistics. , 75* (4), 12–21.

Jacoby, H. (2000). Access to markets and the benefits of rural roads. *Economic Journal , 110*(465), 713-737.

Jacoby, H. G., & Minten, B. (2009). On measuring the benefits of lower transport costs. *Journal of Development Economics, 89*(1), 28-38.

Jalan, J., & Ravallion, M. (2002). Geographic poverty traps? A micro model of consumption growth in rural China. *Journal of Applied Econometrics, 17*(4), 329-346.

Khandker, S. R., Bakht, Z., & Koolwal, G. B. (2009). The Poverty Impact of Rural Roads: Evidence from Bangladesh. " *Economic Development and Cultural Change, 57*(4), 685-722.

Lokshin, M., & Yemtsov, R. (2005). Has rural infrastructure rehabilitation in Georgia helped the poor? *The World Bank Economic Review, 19*(2), 311-333.

MCC. (2012). *Millennium Challenge Account – El Salvador, Monitoring and Evaluation Plan.* Version 6. Retrieved from https://www.mcc.gov/documents/data/SLV_ME_Plan_Sept_2012_MCC_Approved.pdf

Morrison, C. J., & Schwartz, A. E. (1996). State infrastructure and productive performance. . *American Economic Review, 86*(5), 1095–1111.

Mu, R., & van de Walle, D. (2011). Rural roads and local market development in Vietnam . *The Journal of Development Studies, 47*(5), 709-734.

Murray, D. (1998). Sample Size Detectable Difference, and Power. In *Design and Analysis of Group-Randomized Trials* (pp. 349-413). New York and Oxford: Oxford University Press.

Nadiri, M. I., & Mamuneas, T. P. (1994). The effects of public infrastructure and R&D capital on the cost structure and performance of U.S. manufacturing industries. *Review of Economics and Statistics , 76*(1), 22-37.

Norton, S. W. (1992). Transaction costs, telecommunications and the microeconomics of macroeconomic growth. *Economic Development and Cultural Change , 41*(1), 175–196.

Paul, C. J., Ball, V. E., Felthoven, R. G., & Nehring, R. (2001). Public infrastructure impacts on U.S. agricultural production: A state-level panel analysis of costs and netput composition. *Public Finance and Management , 1*(2), 183-213.

Raudenbush, S. W., & al., e. (2011). *Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01) [Software].*

Ravallion, M. (2007). Evaluating anti-poverty programs. *Handbook of development economics , 4*, 3787-3846.

Röller, L., & Waverman, L. (2001). Telecommunications infrastructure and economic growth: A simultaneous approach. *American Economic Review, 91*(4), 909–923.

Rosenbaum, P., & Rubin, D. (1983). The Central Role of Propensity Score in Observational Studies. *Biometrika*, 41-55.

Roy, A. (1951). Some Thoughts on the Distribution of Earnings". *Oxford Economic Papers, 3*, 135-145.

Rubin, D. (1974). Estimating Causal Effects to Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology, 66*, 688-701.

Rubin, D. (1977). Assignment to Treatment Group on the Basis of a Covariate. *Journal of Educational Studies. , 2*, 1-26.

Rubin, D. (1979). Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies. *Journal of the American Statistical Association, 74*, 318-328.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations.* New York: Modern Library.

van de Walle, D. (2009). Impact evaluation of rural road projects. *The Journal of Development Effectiveness, 1*(1), 15-36.

van de Walle, D., & Cratty, D. (2002). *Impact Evaluation of a Rural Road Rehabilitation Project.* World Bank.

Annex 1: Original Sample Design

We assumed a clustered, quasi-randomized evaluation design with treatments administered at the cluster level and with data collection before and after initiation of the treatments. With this design, impact estimates can be measured using the preferred approach of taking difference-in-differences or "double difference": the change in the outcome in the treatment group minus the change in the outcome in the quasi-randomized control (or alternate treatment) group. The purpose of the sample size estimates or power analysis is to determine the minimum detectable impact (Δ) for a given number of sampled clusters (g) and households per cluster (m) in each treatment condition for the evaluation sample.[16] If the impact of the treatment is at least as large as Δ, we will be able to detect it at the 95 percent confidence level with the assurance that at least 80 percent of the time that the null hypothesis of no impact is false (i.e. there is an impact), we would reject this null hypothesis if we have a sample of total size $2mg$. If the treatment impact is less than Δ, we are less likely to detect it, although detection is still possible.

a. Intra-cluster Correlation

The most controversial issue in sample design is the intra-cluster correlation, so we will proceed to make the calculation procedure explicit. DIGESTYC[17] and the Ministry of Public works of El Salvador provided detailed Geographic Information System (GIS) data regarding the road and the location of all the dwellings of northern El Salvador. The intra-cluster correlation of several variables was calculated from a National household survey, the EHPM 2007. Merging the survey and the GIS data, the cantons through which the road runs were identified and matched to the household survey data. The sample universe is constituted by the set of cantons identified by the census as having dwellings within several access time thresholds to the highway: 10, 20, and 30 minutes.[18] The sub-set of cantons that were included in the household survey constitutes "level 1".

For those cantons that were not included in the EHPM survey, the municipality income/expenditure and time allocation data was imputed. This group plus "level 1" constitutes "level 2". In turn, for those municipalities that were not included in the survey, the department data was imputed. This sub-set plus "level 2" makes up "level 3". For the 10 minute threshold, we recommend the use of level 1; for the 20

[16] In addition to g and m, the minimum detectable impact, Δ, is a function of the variance of the outcome variable, its intra-cluster correlation, and the area of influence of the highway being evaluated. See formulae below

[17] Dirección General de Estadísticas y Censos

[18] The accessibility measure is calculated using Geographic Information System (GIS). Digital information of roads, rivers, and the slope of the terrain is used in order to build a measure of accessibility to the northern highway. The proper combination of this data allows for the calculation of the lowest cost path and surface (based on time) from any household to the nearest point of the northern highway.

This model simulates the time it takes a person to reach the nearest point of the highway, assuming that people prefer to travel via highways, roads, and trails and also considering rivers or lakes as barriers if there is no bridge available. The model applies a cost-weighted distance algorithm in order to calculate the accumulated time required to travel from any location to the destination point. In this case, each household gets a weighted average of the time traveled to the highway.

and 30 minute threshold, we recommend level 2 (canton plus municipality). Level 3 is not recommended because the department data may not be a good approximation of the households near the road.

Several outcome variables were used in the analysis. We will summarize the results in the initial power calculations for overall household income, but the analysis also included hours worked per week; both are divided into wage and non-wage agriculture, and wage and non-wage non-agriculture activities.

b. Scenarios for Variance Calculation

There are three important differences between the proposed sample for evaluation and the EHPM sample, all of which are likely to affect the sample variance in the projected sample relative to that in the EHPM sample. First, we estimated the variance of the primary outcomes in the EHPM using only one round of data collection, rather than two. The variance of the difference between the two measures depends upon the variance of each measure as well as the correlation over time between the two measures. Using these data, we did not know this correlation, so we made assumptions about it, which we detail below. Second, we stratified the sample for the collection of these data in order to both balance the sample and reduce the sampling variance. The reduction in sampling variance will depend upon the variance between strata means; the larger the difference between the average outcomes across strata, the higher the variance reduction. Third, the EHPM measures the variance of outcomes related to different levels of current road access, thus the variance at baseline will be smaller given the assumptions of accessibility we are imposing.

Since the three differences between the proposed surveys and the EHPM will certainly affect the variance of primary outcomes, we experimented with power calculations using several different variance estimates. First and most conservatively, we simply doubled the variance of the EHPM outcomes; doing so assumes that the primary outcome will not be correlated across the two surveys, that each strata will have exactly the same mean outcome, and that the treatment will not affect the variance of the treatment. Second, we reduced the doubled variance by 10 percent to simulate a significant decline in sample variance due to stratification. Third, we simply computed the power calculations using the EHPM variance. Finally, we used the EHPM variance less 10 percent to account for gains from stratification, but also assumed an overtime correlation within clusters of 0.5 and an overtime correlation within households of -0.5. Since we also ignored the previous assertion that the baseline variance in outcomes is likely to be smaller than the EHPM variance, the fourth estimate is likely to be the most realistic and the one we proposed to use.

Table 23 presents the results for the different scenarios assumed at the design stage. Note that we fix α at 5% and β at 20% with a minimum detectable effect Δ of 20%.

Table 23 Number of Clusters per Condition[1] and Total Sample Size[2]

Household Income[3] for Each Scenario[4]

Test Number	Treatment Group	Control Group	Intra-cluster correlation[6]	Scenario 1 Clusters per condition	Scenario 1 Total sample size	Scenario 2 Clusters per condition	Scenario 2 Total sample size	Scenario 3 Clusters per condition	Scenario 3 Total sample size	Scenario 4 Clusters per condition	Scenario 4 Total sample size
[5]m=25											
1	T2	T1	0	20	1021	18	919	10	510	10	478
2	T2	T3	0.04	68	3405	61	3064	34	1702	17	851
3	T3	T5	0.06	88	4407	79	3966	44	2203	18	888
4	T4	T5	0.04	112	5586	101	5027	56	2793	28	1385
m=35											
1	T2	T1	0	14	1003	13	903	7	502	7	470
2	T2	T3	0.04	58	4026	52	3623	29	2013	12	836
3	T3	T5	0.06	77	5421	70	4879	39	2711	12	873
4	T4	T5	0.04	94	6612	85	5951	47	3306	19	1361
m=45											
1	T2	T1	0	11	994	10	894	6	497	5	466
2	T2	T3	0.04	52	4661	47	4195	26	2331	9	828
3	T3	T5	0.06	72	6450	65	5805	36	3225	10	865
4	T4	T5	0.04	85	7662	77	6896	43	3831	15	1348

1 The conditions are “treatment” and “control”. The number of clusters in each condition is equal

2 Total sample size (treatment + control)

3 The outcome variable is total monthly household income

4 For the formulae, see text and Murray (1998, chapter 9)

5 Number of observations (households) per cluster

6 Observed in the EHPM with “level 2” households.

7 Parameters: α=0.05; $\boldsymbol{\beta}$=0.20; Δ=0.20