Impact Evaluation of Infrastructure Investments: The Experience of the Millennium Challenge Corporation

Ariel BenYishay and Rebecca Tunstall*

July 2010

* Millennium Challenge Corporation, 875 15th St. NW, Washington, DC 20005. Email: benyishaya@mcc.gov and tunstallrh@mcc.gov. We are grateful to Jigar Bhatt, Ken Fortson, Celeste Lemrow, and Anne Pizer for the time and insights they shared with us. Thanks also to Andrew Carter and Algerlynn Gill for excellent research assistance. The views represented herein are our own and do not represent those of the Millennium Challenge Corporation nor of any other organization.

Abstract:

In many developing countries, aging or inadequate infrastructure is a binding constraint to economic growth. The Millennium Challenge Corporation (MCC), a US government agency providing development assistance, has committed more than $4 billion to upgrade or rehabilitate roads, ports, electricity, water, sanitation and major irrigation systems in 16 countries between 2004 and 2010. In at least eight of these countries, MCC has developed evaluations that will assess the causal impacts of these investments on a variety of outcomes, including household incomes and consumption. These evaluations primarily rely on difference-in-differences estimation, complemented by random assignment, propensity score matching, GIS models, and regression discontinuity designs. The relatively large number of evaluations (13 in all) and the diversity in their approaches offers a unique opportunity to compare these evaluations in terms of the techniques used, their ability to control for selection bias, and their flexibility under changing implementation plans.

Our paper studies the conditions that led to the design of each evaluation, including differing mechanisms for selecting infrastructure to be upgraded. We compare the propensity score matching approaches used in many of these evaluations, noting key observable characteristics used to match treatment and control communities. We also study the GIS modeling approaches used in four of the roads evaluations. Finally, we review the flexibility of each evaluation design in response to changes in the project implementation plans that arise when there are cost over-runs and/or poor policy performance, there are delays in construction, or there are changes to the roll-out strategy. Several of these evaluations will provide the first rigorous evidence on the impacts of highway or secondary road improvement in developing country contexts. Similarly, a number of evaluations will offer important evidence on the extent to which water and sanitation improvements can raise the income level of households. By incorporating multiple methods, a number of these evaluations will also illustrate whether these methods produce different impact estimates, another notable contribution to the literature.

I. Introduction

In many developing countries, aging or inadequate infrastructure is a binding constraint to economic growth. Weaknesses in transport networks, for example, limit connections between farmers, urban markets and producers, and international markets. Where they exist in good condition, rural roads connect villages and small towns, allowing farmers to deliver produce to local markets and obtain agricultural inputs. These roads also feed into larger secondary roads, on which these goods travel to and from larger urban markets. Access to these broader markets spurs non-agricultural production in rural areas. National highways link these urban centers and markets in neighboring countries, as well as ports and airports used to export products.

Many developing countries' governments, however, have been unable to devote the considerable capital expenditures required to rehabilitate, maintain, or upgrade road networks and other infrastructure to their optimal condition. As a result, many rural roads are impassable for large parts of the year, and the time and vehicle damage incurred in traveling along them make transport prohibitively costly for the areas' farmers. Conditions along secondary and primary roads also force users to travel more slowly than they would otherwise and incur higher costs in operating their vehicles. Other types of infrastructure—including electricity and water systems and major irrigation canals—are also frequently inadequate, limiting the extent of output that an economy can generate. To ease these constraints, the Millennium Challenge Corporation (MCC), a US government agency providing development assistance, has partnered with 16 developing countries to support infrastructure improvements these countries have identified as crucial for economic growth. Since its creation in 2004, MCC has committed more than $4 billion to improve infrastructure systems in these countries, including $2.3 billion to upgrade or rehabilitate primary, secondary and rural roads in 14 countries. Along with these capital investments, MCC has also recommended—and in some cases required—that recipient country governments enact new arrangements for funding ongoing maintenance of their infrastructure.

Until recently, many of the claims about the impacts of infrastructure improvement on economic growth were not substantiated with empirical evidence. Advances in program evaluation in the past 15 years, however, have allowed investigators to study these impacts with some confidence in their attribution. This development has been due in part to the broadening use of difference-in-difference (DD) estimation and propensity score matching (PSM) to control for selection bias in non-experimental data (sparked by Rosenbaum and Rubin 1983, Heckman et al 1997, 1998a, 1998b, and Dehejia and Wahba 1999). These methods rely on controlling for pre-intervention differences between treatment and comparison groups and so have been bolstered by the growing availability of pre-intervention data.

Despite these advances and the resulting spate of work on the impacts of infrastructure in developing countries, major gaps in our understanding of these impacts and their mechanisms remain. While the impacts of rural roads on households' material well-being have been studied in a variety of settings, almost no causal evidence exists on the impacts of national highways and major secondary roads on the surrounding populations. Similarly, while the effects of improved water and sanitation access on child mortality has been well-studied, impacts of these improvements on households' near-term income generation and consumption patterns are yet to be documented. More broadly, the behavioral response of households to improved access to electricity, water, sanitation, and irrigation in terms of off-farm employment activities and on-farm crop diversification remains an area of much-needed study.

Against this evidentiary background, MCC has launched evaluations that will assess the causal impacts of its infrastructure investments in eight countries. Evaluations of road investments will estimate effects on transport costs, local prices, and/or household and business incomes, while water, sanitation, and electricity evaluations will track household health, education, and time use, as well as productive uses of this infrastructure. Similarly, an evaluation of irrigation investments in Armenia will estimate shifts in crop compositions and the associated gains in agricultural productivity and total household income as a result of these investments.

MCC's roads evaluations study the impacts of improvements along roads of varying type, length, pre-project condition, and geographic region. The evaluations themselves include both some notable similarities and differences. All of them rely on difference-in-differences estimation, often complemented by propensity score matching. As such, they include a comparison group for the communities that are treated with improved roads, while controlling for both observed differences between the treatment and comparison groups and time invariant unobservable differences. In addition, several of the evaluations include continuous treatment estimators that consider the degree of improvement in travel time and costs as the treatment that communities experience, and relate the intensity of this treatment to gains in household well-being. GIS data will also play a key role in a number of these evaluations, integrating into the matching of treatment and comparison groups, the measurement of continuous treatment effects, and the estimation of the network effects of improving individual segment. Finally, several evaluations are attempting to use thresholds in the selection of segments or the timing of their improvements to identify their impacts.

In a similar vein, evaluations of MCC investments in electricity distribution, water and sanitation systems, and major irrigation canals in five countries are rigorously estimating counterfactual scenarios to capture the impact of these investments on household well-being. These evaluations similarly rely on difference-in-difference estimations, often couple with matching of communities in treatment and comparison areas. In several cases—rural water in Lesotho and electrification in El Salvador and Tanzania—the evaluation involves random assignment of communities and/or households into the treatment group. Looking at macro effects of infrastructure improvement, the evaluation of Georgia's Regional Infrastructure Development activity complements household and firm-level analysis with a computable general equilibrium (CGE) model estimating effects on national output, prices, employment and expenditures.

The relatively large number of MCC infrastructure evaluations and the caliber of and diversity in their approaches offers a unique opportunity to compare these evaluations in terms of the techniques used, their ability to control for selection bias, and their flexibility under changing implementation plans. This paper studies the conditions that led to the design of these evaluations, noting how MCC's pre-investment analysis process sometimes precludes randomized evaluations but can facilitate credible evaluation designs. We assess the PSM and GIS approaches used in the evaluations, highlighting how the combination of these approaches will enable the evaluations to identify the impacts of highway and secondary road improvements in developing countries. These results will provide important information for developing country governments, donor agencies, and individual stakeholders as they consider alternative investments of scarce resources.

We also compare the outcomes studied in each evaluation, noting that some evaluations will identify effects on the prices of agricultural goods, inputs, and consumption goods, while most will also capture effects on agricultural and non-farm production and labor income. Taken together, the evaluations will illustrate the conditions under which different effects of infrastructure improvements are likely to emerge.

Finally, our paper reviews the flexibility of each evaluation design in response to changes in the implementation plan of each project that arise when there are cost over-runs and/or poor policy performance, there are delays in construction, or there are changes to the roll-out strategy.

While we largely focus on evaluations of MCC road improvement investments which make up a substantial portion of the overall infrastructure evaluation portfolio, we note how these differ from MCC evaluations of other types of infrastructure.

The paper is organized as follows: In Section II, we discuss the existing literature studying the effects of infrastructure improvement on material well-being in developing countries. In Section III, we discuss MCC's infrastructure investments and its impact evaluations of these investments. In Section IV, we describe the designs of MCC's evaluations, delving into the PSM and GIS methods used in a number of these evaluations in Sections V and VII. In Section VIII, we compare the outcomes studied in the roads evaluations, while Section VIIII investigates the flexibility of these evaluations in response to project implementation changes. Finally, we offer recommendations for future evaluations in Section IX.

II. Existing Literature

As previously noted, advances in program evaluation in the past 15 years have added substantially to the body of knowledge on infrastructure improvement impacts. Evidence on improvements in road systems has been particularly strengthened. Estache (2010) provides a useful review of these studies, which have primarily focused on rural roads. Repeated surveys of the population around these roads allowed evaluators to compare the changes in costs and production levels around both improved and non-improved roads, controlling for any time-invariant differences between these roads. Matching improved roads to statistically comparable non-improved roads further allowed evaluators to control for differential time trends that were correlated with observable characteristics of these roads. Van de Walle and Cratty (2002) provide an important early example of how PSM could be combined with DD estimators to assess the impact of rural roads in Vietnam. Escobal and Ponce (2004) and Lokshin and Yemtsov (2005) study related issues in Peru and Georgia. Dercon, Gilligan, Hoddinott and Woldehanna (2006) and Khandker, Bakht and Koolwal (2008) instead rely on household-level panel data to control for differences both across and within communities in the likely impact of the road improvements in Ethiopia and Bangladesh. All of these studies find significant effects on household material well-being, with several of the studies pointing to gains in agricultural prices and production, lower input costs, and higher wages as the channels for income or consumption gains.

The mounting evidence in support of the impacts of rural road improvement contrast sharply with the lack of empirical evidence on the effects of highway or secondary road improvements, particularly in developing countries. Evaluators of non-rural roads have struggled to identify a counterfactual that was actually observed in the form of a comparison group. Instead, as Estache (2010) notes, most evaluators have turned to computable general equilibrium models to simulate these outcomes. One notable exception is Garcia-Mila and Montalvo (2007), who use data from geographic information systems (GIS) to separate 20 km segments of Spain's national highway system and find no statistically significant effects on business creation as certain segments are improved and others not. To date, there is little written about the effects of improvements in highways or secondary roads in developing country contexts.

Evidence on the effects other types of infrastructure improvements in developing countries has similarly strengthened significantly in recent years. Access to piped water, for example, has been demonstrated to improve child survival and health outcomes in a number of developing countries (Lee et al (1997), Jalan and Ravallion (2003)). Rauniyar et al (2010) find that Pakistani households in areas that gained access to water and sanitation saw their high school-aged daughters attend school more frequently. The latter authors do not find evidence that households use time savings to increase labor force participation or hours worked. As such, this stream of evidence suggests that water and sanitation improvements can increase the income an economy generates primarily over a longer time horizon, when children's survival and human capital eventually translates into gains in output in their adult years.

Similarly, gains in household access to electricity can lead to additional investments in child welfare, but the degree to which they affect a household's near-term earning potential remains unclear. Chowdhury and Torero (2009) provide some evidence that access to electricity raised Bangladeshi women's labor income, but this additional income was not necessarily compensated with less time spent on unpaid work at home. Thus, while gains in off-farm microenterprise and employment opportunities and on-farm productivity from access to electricity are often posited, these channels are not well documented in the literature.

Finally, irrigation infrastructure improvements might be expected to more directly influence rural incomes by increasing agricultural productivity and enabling transitions toward higher value crops that require significant water inputs. White (2008) presents evidence that improvements in several large irrigation canals in India did raise yields, but not by the expected amount. Moreover, the expected diversification into higher value crops did not take place following these irrigation improvements. As this diversification often determines whether income growth among farmers will be sufficient to justify the large costs associated with irrigation investments, it remains to be seen whether such investments in other countries can yield cost-effective poverty reduction.

III. MCC infrastructure investments and evaluations

MCC was created in January 2004 by the US Congress. MCC provides low- and lower middle-income countries with five-year large-scale grants, called compacts, to fund investments aimed at reducing poverty through sustainable economic growth. Three key principles underlie MCC's work: (1) investing in countries with strong policy performance; (2) supporting country ownership; and (3) focusing on results.

The potential reward of an MCC compact creates a powerful incentive for adopting good policies. MCC only partners with countries that demonstrate certain levels of policy performance relative to their peers. Once a country is selected as eligible for an MCC compact, the country is invited to develop an investment proposal. Maintaining strong policy performance during implementation of the compact is a condition for continued investment by MCC.

The proposal development process is generally led by the eligible country. Eligible countries identify their constraints to achieving sustainable economic growth and poverty reduction and then develop proposals to address these constraints in broad public consultations. Country governments prioritize areas for MCC investment and then MCC works closely with the country to refine the program and agree on the final compact activities. The prioritization of areas for investment by MCC happens differently in each country. In Armenia, for example, the compact proposal was developed through a consultative process in which more than 1,200 individuals participated and some 230 written proposals were received on particular investment projects. Improved physical infrastructure was highlighted as a priority area, so the Government of Armenia selected improvements in irrigation infrastructure and roads as the main areas for MCC support. In the road sector, the original proposal from Armenia to MCC included over 1,100 kilometers of rural roads that had been selected from the Lifeline Network through an initial screening that took into account population, estimated traffic, and distance to the capital, Yerevan. The long list of road links was pared down through due diligence as more information became available. In El Salvador, in contrast, the proposal was based on an existing national plan that had resulted from a process led by the National Commission for Development with broad participation from local governments, private enterprises and civil society. The Northern Zone portion of the plan was selected for MCC investment and it included, among other investments in community infrastructure, education and business development, the Northern Transnational Highway that had been promised to the people of the Northern Zone for over 50 years.

To increase aid effectiveness, MCC has created an analytical framework that systematically includes a strong focus on results, including significant *ex ante* analysis. The compact proposal process starts with technical analysis to identify constraints to growth and possible remedies. Then, benefit-cost analysis (BCA) and distributional analysis are conducted to assess the economic justification of the proposed investments and the distribution of the expected benefits.

The BCA is conducted in the form of an economic rate of return (ERR) model in which the expected benefits of a project are compared to the estimated costs. The benefits are always estimated changes in local real disposable income, although the pathway to achieve those changes in income may differ from country to country depending on the type of project, the local context, and the available data at the time of the analysis. The ERR also requires assumptions about what would have happened without MCC's investment – the counterfactual.

For roads, the most commonly used ERR model is the Highway Development and Management system or HDM-IV and MCC uses this model for most of its road projects. During the design of road improvement projects, the expected changes in road quality (roughness), speed and traffic are forecasted. Those estimates are then used to calculate the expected savings in vehicle operating costs (VOC) and travel time, which are monetized, aggregated and compared to the cost of the initial investment plus ongoing maintenance and repair over the life of the project (typically 20 years). In some cases, the World Bank's Roads Economic Decision model (RED) is used for low-volume roads, as it designed specifically for unpaved roads with high uncertainty of initial roughness and traffic measures.

Even though most road economic analyses start with estimating savings in VOC and travel time, additional benefits have been included in a few of the countries discussed in this paper. Some of those include increases in agriculture production, increases in land values along the road, and income increases because of improved access to health and education facilities. This is always "tricky" territory though because the available evidence on what those other benefits are is lacking. Many countries project large improvements in health and education because of roads when they present their proposals for funding; however, there is little rigorous quantitative evidence available to back those claims up and to justify inclusion of those benefits in *ex ante* ERRs. Current MCC policy is that *ex ante* ERRs are calculated on approximately a large share of project activities, and road sections that fall below the minimum hurdle for an ERR are not commonly included in compact programs.[1]

Similar to road investments, when MCC receives a proposal for other types of infrastructure, economic rates of return are calculated to estimate the potential benefits and compare them to the cost of the project. However, the models used are not as standard as the HDM-IV used for road modeling. The benefit-cost models for other types of infrastructure must take into account the intermediate results of improving the infrastructure and the productivity and behavioral changes that result. For a water and sanitation project, the ERR may include the value of time savings from no longer having to travel long distances and wait in line for water or it may include the benefits of less disease in the form of more days in school or work and lower medical costs. The benefits for an electricity project may include cost savings from no longer having to buy alternative sources of energy, new business development within a newly electrified household or increased study time for children in the evenings. Just as in the roads analysis, in some cases, a portion of the expected benefits do not have a wealth of rigorous studies to pin down their magnitude. Therefore, the models are developed with the best available information at the time and key assumptions are highlighted for future analysis.

During implementation, MCC requires strict monitoring and reporting. Finally, MCC commissions rigorous independent impact evaluations when there are lessons that could be applied to future funding decisions or project design, when there is a need for evidence about a particular intervention, and when they are feasible.

MCC's impact evaluations are designed to analyze the impact of its investments on economic growth and poverty reduction, as this is the mission of MCC. Through these evaluations, then, MCC reports on the initial goals stated in each compact and reports to external stakeholders such as Congress about the impact of the funding.

More specifically, all of MCC's evaluation work is focused on analyzing the outcomes that correspond to the *ex ante* estimates of economic impact used to justify program investments. The assumptions about changes in income and the counterfactual that are made during project development are tested through the impact evaluations and then used to calculate *ex post* ERRs.

In addition, impact evaluations are designed to compile evidence to improve the selection and design of future projects. Therefore, one of the goals of these road impact evaluations is to compile estimates for road impacts that could be included in *ex ante* ERRs for future roads projects to more accurately assess the potential of a project compared to its cost and to inform investment decision-making. MCC is committed to making the results of all of its impact

[1] Hurdle rates for ERRs have historically fallen between 8-15%, depending on the country and its recent growth experience. A standardized hurdle rate for new activities and changes to activities already under implementation is currently set at 10%. All of MCC's economic rate of return models can be found on its web site at: http://www.mcc.gov/err.

evaluations public, so that evaluation results may be used by host-country governments and other donors to improve the targeting of funds.[2]

MCC has signed 20 compacts with partner countries committing $7.2 billion in assistance. Thirty-three percent, or $2.3 billion, has been allocated for transportation activities, the largest sector MCC in which is investing. These road investments comprise 14 projects, and impact evaluation strategies have been developed for half of them. These seven projects—each in a distinct country—represent $1.14 billion and 2,189 kilometers of road improvements. They include all types of road improvements from new roads, to upgrading rural roads and rehabilitating highways.

Table 1: MCC Road Projects with Impact Evaluations

Country	Project	Budget Allocation	Number of Kilometers to be Improved	Evaluation Designer
Armenia	Rural road rehabilitation	$67 million	24 km[3]	Mathematica Policy Research
El Salvador	Secondary road openings and upgrades	$233 million	290 km	Social Impact with the International Food Policy Research Institute
Georgia	Main road rehabilitation	$204 million	120 km	National Opinion Research Center
Ghana	Feeder road rehabilitation	$70 million	357 km	National Opinion Research Center
Honduras	Secondary and tertiary roads Highway	$21 million $119 million	750 km 109 km	National Opinion Research Center
Nicaragua	Main and secondary roads	$58 million	74 km	Millennium Challenge Corporation
Tanzania	Trunk roads and rural roads	$369 million	465 km	Economic Development Initiatives, Ltd

Similarly, MCC's evaluations of non-roads infrastructure projects also cover an array of activities in a number of different countries:

[2] Descriptions of MCC's impact evaluations can be found on its web site at: www.mcc.gov under "Programs and Activities" and then "Impact Evaluation."

[3] The Armenia budget allocation was originally committed to rehabilitate up to 943 kilometers; however, the activity is currently on hold. Twenty-four kilometers is the revised target for number of kilometers to be completed even though the budget allocation has not officially changed.

Table 2: Other MCC Infrastructure Projects with Impact Evaluations

Country	Project	Budget Allocation	Expected Beneficiary Count	Evaluation Designer
Armenia	Irrigated Agriculture Project[4]	146 Million	420,000	Mathematica Policy Research
El Salvador	Rural Electrification Activity	33 Million	235,000	Social Impact with the International Food Policy Research Institute
El Salvador	Water and Sanitation Activity	24 Million	90,000	Social Impact with RTI International
Georgia	Regional Infrastructure Development Activity	57.7 Million		Tbilisi Business Service Center and ACT Consulting
Lesotho	Water Sector Project	164 Million	644,000	National Opinion Research Center
Tanzania	Electricity: Distribution Rehab. & Extension	206.47 Million	1,080,000	Mathematica Policy Research

To date, none of the MCC infrastructure projects considered in this study has been completed as originally planned; however, the Honduras road project, which was one of the first compacts to be signed, will finish this year. Table 3 displays the state of implementation of the projects in these 8 countries.

[4] This project includes activities in various sectors. The evaluation discussed in this paper is covering only a portion of the infrastructure activities – the tertiary canal rehabilitation.

Table 3: State of Implementation as of March 2010

Country	Compact Begin Date	Compact End Date	Roads Completion		Construction Progress on Other Infra. Activities
			Percent Complete[1]	Kilometers Complete	
Armenia	Sept. 2006	Sept. 2011	100%	24 km	7%
El Salvador	Sept. 2007	Sept. 2012	32%	0 km	0% Water 8% Elec.
Georgia	April 2006	April 2011	48%	0 km	20%
Ghana	Feb. 2007	Feb. 2012	12%	0 km	n/a
Honduras	Sept. 2005	Sept. 2010	43%	310 km	n/a
Lesotho	Sept. 2008	Sept. 2013	n/a	n/a	0%[2]
Nicaragua	May 2006	May 2011	99%	74 km	n/a
Tanzania	Sept. 2008	Sept. 2013	0%	0 km	0%[3]

[1] Percent complete is measured as the amount disbursed divided by the total amount contracted for construction works

[2] Urban Water: Design work for distribution systems continues this quarter. Rural Water: The procurement process has begun for 25% of the rural water systems and latrines.

[3] In Tanzania, design, environmental and social preparatory work associated with the upgrade of transmission systems and distribution line extensions in six underserved regions of the country continues.

One can reasonably ask how these projects were selected for evaluation from MCC's broader portfolio of infrastructure investments. The investments being studied in these impact evaluations total approximately more than $1.5 billion, or slightly more than a third of MCC's overall infrastructure investments. In determining how rigorous an evaluation of a given project should aim to be, MCC considers several factors. Technical feasibility is always a prerequisite, but learning potential, need for evidence justifying the program, and cost also play an important role in evaluation selection.

The share of infrastructure investments with existing impact evaluations may be slightly misleading because a number of these investments are still quite early in their implementation and may yet have impact evaluations, as may be the case with road improvement projects in Burkina Faso, Moldova, and Senegal, for example. Once the evaluations of these recently launched projects are designed, we expect that the share of MCC infrastructure funding covered by impact evaluations may rise above 50%.

Moreover, even in cases where rigorous statements of causal impacts cannot be made, MCC is devoting significant resources to evaluations of the outputs and outcomes of these activities. For example, for a number of activities, comparisons of conditions before and after the investments, although made without the observation of a comparison group, can yield helpful evidence on the bounds for the program's potential effects. Evaluations of water sector investments and upgrades to Mafia airport in Tanzania are examples of such before-and-after studies. In other cases, independent organizations will be engaged to use data collected through MCC's project administration and M&E systems and through post-project surveys to provide objective assessments of the potential range of impacts given evidence on program efficiency and effects among the treated populations.

Finally, it is worth noting that the impact evaluations of these investments can be designed to answer two questions. The first is "what is the impact of MCC's investment in road improvements?" An evaluation designed to answer this question will model the counterfactual and estimate the impact of MCC's funding. If road improvements, for example, are made on any of the comparison roads, then that would reduce the measured impact and imply that those roads may have been improved even without MCC's funding. In some instances, the fungibility of public finances means that MCC's funding may actually enable some of the improvements along comparison roads, which raises concerns about contamination of the comparison group. The second question that the evaluation can be designed to analyze is "what is the impact of infrastructure improvements, regardless of the funding source?" An evaluation designed to answer this question would include all improved infrastructure in the treatment group even if it had been improved by another funding source such as another donor or the host country government. The impact evaluations discussed herein contain a mix of these two approaches, offering an important contribution to the existing literature.

IV. Evaluation Designs

MCC compacts in the eight countries considered in this paper are funding improvements of infrastructure of varying type, scope, and pre-project condition. In all of these compacts, the selection of projects for investment was the result of a process that explicitly considered the likely economic impacts of rehabilitating or upgrading each piece of infrastructure, as well as other features of the infrastructure that are likely to be correlated with outcomes. As described in Section III, the project selection process involved public consultations in the country, prioritization by the government, and benefit-cost and distributional analysis of proposed investments. We discuss these in greater detail below and note how evaluation designs must take each of these selection features into account.

Because road investments are a broad treatment and MCC's consultative and analytical process frequently leads to selection of individual road segments, random assignment of individual, communities, or road segments to treatment and comparison groups is frequently not possible. To nonetheless generate a comparable group and attribute impacts to MCC's road investments, all of the organization's roads evaluations utilize PSM and DD. In addition, several evaluations are using continuous treatment methods that measure the actual changes in transportation access for different communities and compare communities for whom these changes are greater to those where changes are smaller. Finally, several evaluations have attempted to use regression discontinuity methods based on a number of thresholds in the selection and roll-out of road segments, with varying success.

Investments in extending the electricity and water and sanitation distribution systems can be more targeted, as they require connection by households or businesses to these systems. In several cases, MCC evaluations involve random assignment of communities, clusters, or households to treatment groups receiving earlier access to improved water systems and coupons to facilitate connections to the electric grid. In other cases, regression discontinuity methods using threshold levels of costs/pricing are being attempted, meeting varying degrees of success (much as in the roads evaluations).

Table 4 details the methods being used in MCC's infrastructure impact evaluation.

Table 4: Evaluation methods

Country	Project	DD	Randomized assignment	PSM	Continuous Treatment	RD	CGE
Armenia	Irrigation	Y					
Armenia	Roads	Y					
El Salvador	Electricity	Y	Y	Y			
El Salvador	Roads	Y		Y		Y	
El Salvador	Water & San.	Y		Y			
Georgia	Regional Infrastructure	Y					Y
Georgia	Roads	Y			Y		
Ghana	Roads	Y		Y	Y		
Honduras	Roads	Y		Y	Y		
Lesotho	Rural Water	Y	Y				
Lesotho	Urban Water	Y				Y	
Nicaragua	Roads	Y		Y			
Tanzania	Roads	Y		Y			
Tanzania	Electricity	Y	Y	Y		Y	

For roads investments, the consultative and analytical basis for funding some individual road segments and not others precludes the use of randomization for most of MCC's road improvement project evaluations. In other types of projects, the selection process may determine interventions that are feasible—say, training on intercropping for maize farmers in Malawi—but may not determine the optimal allocation of these interventions across observably homogeneous communities or individuals. That is, the pool of people who are eligible to receive the treatment and for whom the impact of the treatment is indistinguishable *ex ante* is sufficiently large that the treatment can be randomly allocated among them with no effect on the *ex ante* estimates of its returns. For example, a farmer training program may be designed to inform maize farmers about the benefits of intercropping, but whether these effects will be larger for farmers in certain communities may not be well known. The program may therefore be randomly assigned within the pool of otherwise homogeneous communities.

Road improvement, on the other hand, is a rather broad treatment—an intervention that is not targeted to individuals or communities, but to stretches of road. One cannot randomly assign treatment and control communities along the same segment selected for improvement. If one conducts *ex ante* ERR analysis on individual road segments as the basis for selection, all communities along these segments would be intended for treatment. The level of analysis and the level of treatment coincide entirely, meaning allocation of the program based on the *ex ante* analysis precludes random allocation of the program among lower level units. Given MCC's focus on *ex ante* analysis at the lowest unit for which confident estimates are possible, it is not surprising that few road improvement evaluations involve randomization. In fact, in only one case—Honduras' rural roads program—was randomization a realistic option. In this case, there was not sufficient data available to estimate the segment-by-segment returns accurately, making randomization among the potential segments a possibility. Although a randomization was initially considered, implementation plans eventually ruled out his possibility.

If a country's initial proposal includes a set of rural road segments, the ERR from investment in each segment is estimated, and a subset of segments for which this rate exceeds a pre-determined hurdle rate may be recommended for funding. The variation in the flow and composition of existing traffic and the population base along the road segments is the primary determinant of the expected rates of return for improving each segment. This selection process thus links the existing conditions, traffic, and population base of each road segment to its probability of being selected for investment. Of course, the appraisal process produces estimates that have a good deal of uncertainty of the true gains from improving each road segment. Nonetheless, if the noise in these estimates is not too great, the set of road segments recommended for improvement will exhibit significant positive selection.

When a large pool of road segments is considered for potential improvements, the ERR-based selection process provides a natural discontinuity that can be exploited to evaluate the effects of these improvements. Segments with ERRs that are just above the hurdle rate may not differ significantly from those just below the rate, except in their probability of investment. The most rigorous regression discontinuity (RD) approach would limit the sample to those roads near the hurdle rate, and make the aforementioned comparison between those just above and just below the hurdle rate. Unfortunately, in none of MCC's impact evaluations was the pool of potential road segments so large that this approach would yield a large enough sample to precisely detect the project's impacts. The evaluation of Armenia's Rural Road Rehabilitation Project (RRRP) offers a useful example of an alternative approach. The evaluation was designed to include the full range of road segments considered for rehabilitation in the sample, while controlling for a smooth function of the pre-investment ERR (Fortson and Rangarajan 2008). If the relationship between the ERR and the outcome variables studied in the evaluation exhibit a stable relationship that is accurately modeled in the estimation, the sharp discontinuity in the ERR at the hurdle rate will identify the program's impacts.

In the case of Armenia's RRRP, road segments were initially divided into three packages of potential road segments that would be rehabilitated. The intention prior to the launch of construction was to select roads from each package using the ERR hurdle rate criteria. All of the road segments in the first package cleared the ERR hurdle rate, making the RD design dependent on the second and third packages to generate a control group of segments. However, the subsequent appreciation of the Armenian Dram caused the costs of rehabilitation to rise substantially. At the new costs, only the road segments in the first package could be funded. Rather than raise the hurdle rate and fund a smaller share of roads from each package, the most efficient project re-design involved funding only the first package of roads. As we discuss in Section VIII, modifications to the RD approach and alternative designs were also considered in response. More generally, the validity of the RD design relies on the consistent use of an explicit analytical framework for road segment selection—a practice that has been challenging to follow even among donors who mandate it as part of their "best practices".

When road segment selection does not follow such an explicit analytical framework, or when the pool of potential segments is relatively small, other methods must be used to estimate the counterfactual. This is particularly true of improvements to national highways or other major roads, for which the number of potential segments for improvement may be limited. For these situations, DD estimators could control for differences in initial conditions across road segments. DD estimation compares the *changes* in outcome variables along treated roads to those along comparison roads. Since most of the parameters used in the HDM-IV and RED models vary across road segments only in terms of their initial characteristics, if these models were the sole basis for selection from a known pool of potential segments, DD estimators would effectively control for selection bias. However, these conditions are frequently not met. For example, a country eligible for an MCC compact may propose only a small set of segments for funding, with all or most of these segments appraised as having ERRs above the hurdle rate. For this and other reasons, one cannot confidently say that the ERR model parameters were the sole basis for the selection. Without explicit comparison to alternative segments that were not selected, the possibility of selection bias remains a key issue.

To overcome this bias, evaluations are increasingly turning to PSM. As previously noted, Van de Walle and Cratty (2002) provide a seminal example of how PSM, in combination with DD, can be used to control for differences between treatment and comparison road segments based on their observable characteristics. Taken together, PSM and DD limit the effect of selection bias to the effect of unobserved characteristics on differential trends across segments. This approach is used in all seven MCC evaluations considered herein (although it was originally conceived as a "back-up" alternative in several of these evaluations). The core elements of this approach require panel data married with observing a broad range of initial characteristics on which to match units. In the next section, we discuss the varying data and matching process used across the seven MCC evaluations.

Four of these evaluations (El Salvador, Georgia, Ghana, and Honduras) utilize GIS to generate data on physical and agricultural conditions along road segments and incorporate this information into the PSM. For example, in Honduras, GIS data allows roads to be matched in terms of their elevation and the surrounding areas' soil type, soil fertility, and rainfall (in addition to other observables). Where it is regularly updated, GIS information can also capture improvements in comparison roads that are funded by other non-MCC projects (including domestic roads rehabilitation programs). Importantly, GIS information can be used to model network spillovers that are likely to occur from improvements along specific segments. This integration of GIS into these evaluations is relatively unique and represents a real contribution to the literature. In Section VI, we highlight how this practice has varied across these four evaluations and what we are likely to learn about this practice based on results of these studies.

In addition to the traditional delineation of communities into treatment and comparison categories, a number of the evaluations also consider treatment as a continuous variable based on the changes in accessibility that occur when a nearby road is improved. The recently developed methods analyzing continuous treatment (Imbens 2000, Behrman, Cheng and Todd 2004, Hirano and Imbens 2004) can be likened to a "dose-response" framework in the medical field. Evaluations of MCC-funded improvements in El Salvador, Georgia, Ghana and Honduras will assess the impact of incremental reductions in travel costs (both time and direct costs) to the nearest market, major towns, or a weighted composite of destinations.[5] In this framework, it is the intensity of treatment that varies,

[5] The Tanzania evaluation will not consider treatment as a continuous variable, but communities have been categorized in terms of their travel time to a treatment/comparison road. The difference in outcomes for communities near an improved road (vis-à-vis communities near unimproved roads) can be compared to the difference in outcomes for communities farther from these improved roads (vis-à-vis outcomes for

and there is no formal control group (no group receives zero treatment). The average dose-response function characterizes the mean impact estimate per unit reduction in travel costs, which can be multiplied by the mean reduction experienced in the sample to obtain the mean total impact. This offers a cross-validation of the PSM-DD results that will be obtained in these evaluations.

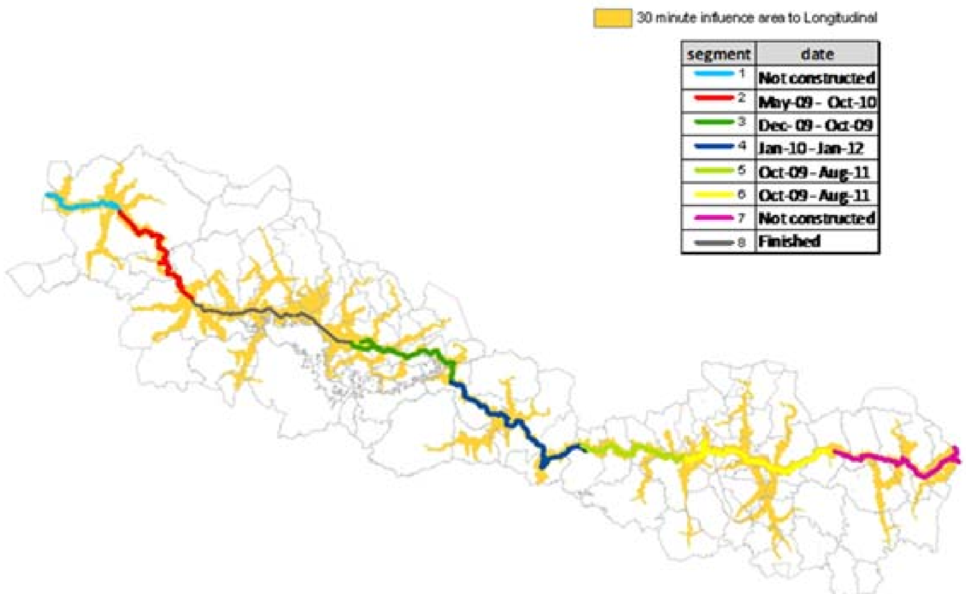
Estimating travel costs in this framework is most carefully accomplished by considering costs reported by households or communities, estimates derived from GIS pathway models, and additional cost data from transport operators. Several of the evaluations utilize primarily GIS-based data, in which the characteristics of each road are considered in terms of the operating costs for vehicle owners (say, the "wear and tear" and fuel costs of driving along a hilly, circuitous, unimproved road). The El Salvador evaluation calibrates these costs using data from domestic transport companies. Other road evaluations calculate an "accessibility index" that converts the minimum-cost path to predetermined destinations onto a normalized scale. Generally, such continuous measures therefore provide robust estimates that can be used to estimate continuous treatment effects.

Finally, one road evaluation has also attempted to utilize the variation in timing of construction along different segments to identify the impacts of these improvements. In El Salvador, the construction of 5 road segments has followed a timeline that includes distinct dates of completion for each segment (see

Figure 1). The evaluation will thus use an RD design, where the thresholds for treatment are the dates of completion of segment. Controlling for a smooth function of time trends, this approach compares earlier implemented segments to later completed segments, with the assumption that the timing of construction was not selected based on the differences in the likely impacts of the construction (i.e., upgrading roads with higher existing traffic first, etc.). To the extent that the timing of construction is driven by engineering requirements or contracting variation that are different from historical features of the areas around each road segment, this approach can provide valid estimates of impact.

communities at a similar distance from unimproved roads). This difference-in-difference-in-difference approach can be likened to a more continuous treatment.

Figure 1: El Salvador Road Improvement Timeline



Reproduced from Torero (2009), page 9.

While treatment of a population with improvement in roads can be fairly broad or continuous, treatment with access to other types of infrastructure is often more discrete. Households access water, sanitation, electricity, and irrigation services via discrete connections, and these connections can vary within communities. These types of infrastructure offer additional opportunities for evaluation designs that are not feasible for roads improvements. MCC's evaluation of extensions of the electricity grids in El Salvador and Tanzania involve random assignment of coupons to facilitate household connections to the extended grids. MCC's rural water supply project in Lesotho involves a random assignment of the timing of construction of 100 small rural water systems (bore holes, gravity systems, and other technologies). Meanwhile, an evaluation of improvements in access to water and sanitation in urban Lesotho is using discontinuities in the pricing of connections to the main lines based on thresholds in the distance between a residence and the main lines.

At the same time, evaluation designs for rural roads and other types of infrastructure do not differ substantially in every case. Evaluations of water and sanitation improvements in rural El Salvador, investments in major irrigation canals in Armenia, and regional infrastructure development in Georgia are being carried out using matching and DD techniques similar to those used in evaluating many roads projects. The latter evaluation in Georgia will complement the household and firm-level analysis with a CGE model estimating the effects of the improved infrastructure on national-level output, prices, employment, and expenditures.

The randomized assignment approaches used in Lesotho, El Salvador, and Tanzania are particularly noteworthy because they will offer some of the most rigorous evidence on the effects of rural water and electricity supply in very poor countries. In Lesotho, the Department of Rural Water Supply (DRWS) identified 250 small community water systems (each serving 100-180 people) that would be fully designed and, if feasible, implemented under the MCC compact. 100 of these systems were designed by August of 2008, and these were randomly assigned into early and late construction groups. These 100 systems were dispersed across 10 districts, with each district having exactly 10 systems. In August of 2008, a public randomization was held in which 5 of the 10 eligible systems in each district were selected for implementation in the first year of the project, with the remaining systems slated for implementation in the second year. The evaluation will thus identify the effects of the improved water systems by the differential timing of changes in outcomes among treatment ("early") and comparison ("late") groups.

In El Salvador and Tanzania, random assignment of coupons to facilitate connection to the electric grid among subsamples will supplement PSM-based estimates across the full evaluation samples. Both projects will involve extending new distribution lines to communities, with households in these communities facing costs to connect to the new lines. The coupons can therefore be considered a randomized encouragement design, in which some households would have connected to the new lines even in the absence of the coupons and some won't connect even if offered the coupons. The effects of electrification will thus be identified by the share of households that connect to the grid only because of the coupons.

This random assignment is a crucial cross-validation of the PSM approaches taken in the broader sample, about which significant concerns remain. One challenge is that a major characteristic determining propensity of connection to the newly extended grid may be the distance of a household from the new lines. In comparison communities, however, the location of those lines is unknown as-of-yet, and thus identifying households that would be close to the hypothetical lines and offer good comparisons is challenging. This is problematic if the location of the lines is correlated with other unobserved geographic features, which may or may not be major sources of bias.

Finally, thresholds in costs or prices associated with extending the electric grid and offering water and sanitation connections can also serve as useful RD-based identification strategies. In Lesotho, the Urban Water and Sewarage Authority (WASA) uses a "distance from the main" standard to determine the connection fee for a particular residence or business, with a step scale fee structure. A particularly crucial cutoff takes place at the distance of 100 meters from the main, with households just a meter closer to the main paying 3,000 *maloti* and those just a meter farther paying 4,000 *maloti*. The distribution of households around the 100 meter mark provides a reasonably large, balanced sample over which to estimate the program effects.

In Tanzania's energy evaluation, however, attempts by evaluators to retrofit an RD design based on an imputed threshold rule for subproject funding were not successful. Evaluators explored the degree to which subproject funding decisions were based on a threshold level of projected revenue/capital expenditure ratios for subproject locations, finding that only 13 out 240 subprojects were both below the threshold level and not funded. Evaluators therefore turned to PSM and randomized assignment of coupons to identify the program's effects.

One additional feature of several of the evaluations considered herein is their use of multiple methodologies to identify the program impacts. These evaluations are particularly informative because they can be used to compare and cross-validate the methodologies. Taken together, this set of evaluations can provide a sense of what the likely results may have been of using alternative approaches in some of these evaluations. For example, if (hypothetically) Georgia's continuous treatment estimates show effects that are 20% lower than those based on PSM-DD estimates, this might suggest that Nicaragua's PSM-DD estimates may overstate the program's effects.

V. Propensity Score Matching

In combination with a DD approach, the use of PSM is intended to control for differences in trends between communities. That is, some communities may experience faster growth in agricultural productivity, gains in land values, and diversification of products available in local markets. These changes may be closely related to the reasons that infrastructure in or near these communities is selected for rehabilitation or upgrading. For example, some communities may be politically represented by parliamentarians who are very successful at obtaining funding for their districts for both road improvements and other key investments in, say, land tenure systems or agricultural extension services. If such correlated investments are made over similar timeframes, the selection of treatment communities will be spuriously correlated with trends in outcomes of interest. Controlling for such cross-sectional differences thus remains crucial even when DD estimators are used.

In many studies conducted using existing samples, these cross-sectional differences must simply be controlled for using observed characteristics. However, in studies designed with such considerations explicitly in mind, investigators can directly consider these differences when drawing their samples. The primary benefit of doing so is that one can select these treatment and comparison community samples to be as similar as possible in their observable characteristics. PSM provides a statistic that summarizes the comparability of these two samples and thus allows one to draw samples that maximize this comparability.

Of course, this approach depends not only on what characteristics are observable but which characteristics are actually observed. Observing these characteristics generally means using secondary data from existing government or other sources and adding to that primary information gathered through pre-intervention surveys. There is a core tension, however, between observing as much information as possible prior to conducting a PSM exercise and maintaining a large analytical sample. If matching is done *after* a baseline survey, many communities covered by the survey may subsequently be deemed to be poor comparisons and dropped from the analytical sample. Thus, most of the MCC evaluations attempted to gather as much information as possible from existing sources.

The Tanzania roads evaluation is using a modified approach. Using existing Government of Tanzania data, comparison roads were first selected to ensure that they would be in the same district and agro-climatic zone and would be of a similar initial quality as the treatment roads to which they are to be compared. In most cases, a comparison road is also within the same ward and electoral constituency as the treatment road. Once this sample of roads was complete, villages along each road were randomly sampled, and sub-village neighborhoods (known locally as *vitogoji*) were stratified as being near or far from each road, based on whether the travel time by foot to the road is < 30 minutes ("near") or >= 30 minutes ("far"). The full community and household pre-intervention surveys were then administered in these *vitongoji*. These *vitongoji* were then matched using PSM on the basis of this data (see below). Because the intervention roads run through five districts with varying features, the matching was first conducted district by district. Comparing the propensity score distributions of treatment and comparison *vitongoji* highlights the challenge in finding a common support (ensuring that only *vitongoji* with propensity scores that are similar to those of the other (treatment/comparison) group are included in the estimation). As Imbens and Wooldridge (2008) highlight, neglecting this common support can be a source of substantial bias in PSM estimates. In the case of Tanzania's district-by-district estimates, a substantial number of comparison *vitongoji* were dropped from the sample because their propensity scores were lower than any of those exhibited by treatment *vitongoji*. To preserve a larger sample, the matching was conducted for the whole sample jointly, with district fixed effects controlling for district-level unobserved differences in mean outcomes. Figure 2 shows the distributions for these alternative specifications. The full-sample matching weakens the ability of the evaluation to claim that all district-level effects have been controlled for through the matching, as district-level heterogeneity may still be present in the responsiveness of outcomes to the literacy rates or other covariates in a *vitongoji*. Nonetheless, this multi-stage approach allowed for better matching using a larger set of observables than would have been the case had the matching been done prior to the baseline surveys and had only secondary data from the

Government of Tanzania been used. At the same time, the possibility that a greater share of the sample may be dropped because of the lack of common support remains an important consideration for future evaluations.

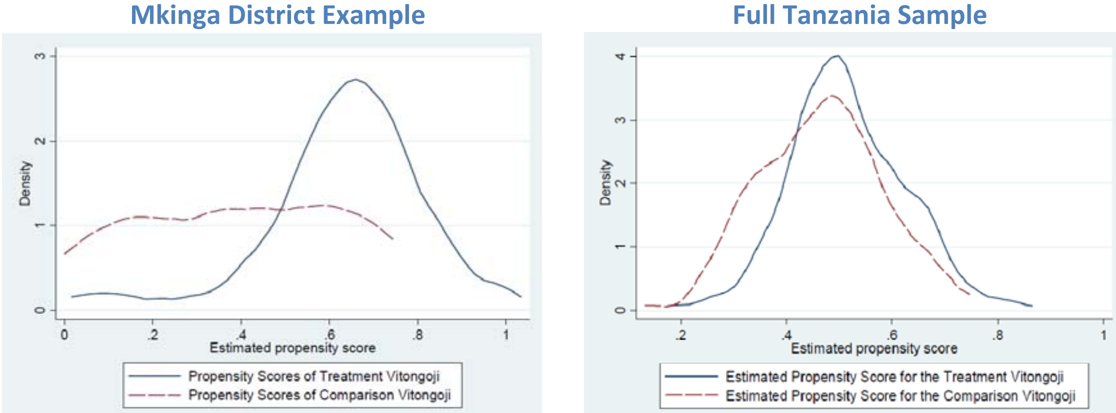
Table 4: Propensity Score Matching for Full Tanzania Sample

	PS1	PS2
Distance to Road	-0.005	-0.005
	(0.003)	(0.003)
Crop Income	0.548	0.303
	(0.805)	(0.840)
Literacy rate	-0.705	-0.668
	(0.763)	(0.771)
Poverty Headcount	-0.059	-0.127
	(0.368)	(0.373)
% Moved up LoL	-0.321	-0.341
	(0.393)	(0.400)
% Female-Headed Households		-3.663
		(3.005)
TLU		-0.116
		(0.090)
District Dummies	Yes	yes
Pseudo R Squared	0.03	0.04

Reproduced from EDI (2009), page 85. Standard errors in parentheses. LoL is the “Ladder of Life” survey instrument used to capture individuals’ general well-being. TLU is Tropical Livestock Units, a count of the livestock each household owns, normalized by FAO weights.

Figure 2: Propensity Score Distributions, Mkinga District and Full Tanzania Sample

Mkinga District Example | **Full Tanzania Sample**



Reproduced from EDI (2009).

One other feature of matching roads or communities within relatively narrow boundaries (like wards) may be that spillovers from treatment roads/communities may affect comparison groups. For example, a road improvement in one part of the ward may reduce the effective price of agricultural produce from that part of the ward at the market located at this ward. One result may be that a comparison community located in this ward struggles to compete with these lower prices and sells less of its produce at this market. If the evaluation compares changes in agricultural production in these two parts of the ward, it would suffer from comparison group contamination. One thus needs to carefully balance the higher quality of matching based on geographical closeness and the possibility of spillovers threatening the validity of the evaluation.

VI. Geographic Information Systems

As previously discussed, evaluations of road improvements can be particularly strengthened by integrating GIS data into their designs. GIS information can be useful in improving the matching process, in estimating changes in road quality due to other programs, and in estimating the network effects of improvements in individual segments. GIS information on roads is generally derived from maps and databases maintained by national roads authorities, satellite imagery, and census/surveys information. This information can also be merged with other national data on natural resources and other infrastructure to create a multi-layered database that allows one to better characterize each community's relationship to these features. Naturally, the ability to use GIS data in evaluations depends on the amount of such data that exists within each country, particularly data on the national road network maintained by the national roads authority. Also important is the regular updating of this database to reflect changes in conditions over the course of the project, both along roads targeted for improvement and those outside the project's scope.

GIS can be used to strengthen the matching process by more precisely measuring the geo-physical and road conditions each community experiences. One important set of these features are agricultural factors, including soil quality, average rainfall, and elevation, which are likely to be correlated with the impact of transport improvements on agricultural production in the area. Information on each road segment is also particularly important, including the quality of the road, the volume of existing traffic, and the amount of congestion regularly experienced on it. In addition, GIS information can be used to calculate the transport costs to types of destinations. In El Salvador, for example, the targeted destination was the nearest town/city with 25,000 or more inhabitants. The minimum-cost route taking into account travel time and road type was identified for each community, and communities can be matched using this factor as one of the variables in the propensity score estimation.

Because GIS data is generally national in scope, it can be used to calculate transport costs for all communities in the country. Thus, this data can allow one to estimate the network effects of road improvements. That is, one can estimate the changes in outcomes for out-of-sample communities for which the minimum-cost pathways run through segments improved by a given project. This is most crucial when the improvements affect major national highways, which are likely to affect many communities in the country. In the Honduras evaluation, for example, the GIS model implemented by National Opinion Research Center (NORC) found that the improvement of the CA-5 highway will likely benefit every village in the country, if all major potential destinations are considered (including cities and ports). By using the continuous treatment methodology to estimate the outcomes per unit reduction in travel costs, one can then impute the likely outcomes that would occur in out-of-sample communities based on their calculated reductions in travel costs. These estimates can provide a more complete sense of improvements in national infrastructure, particularly when they are based on updated data on road quality and usage throughout the network. In Georgia, for example, origin-destination studies are expected to provide context for estimates of these national-level spillovers by detailing the volume and paths of traffic flows after the improvement of the Samtskhe-Javakheti road. More generally, validating the changes in travel costs along the entire network with survey responses or global positioning system data and merging these with outcomes estimated in the evaluation sample can provide estimates of national-level impacts from major road improvements. In essence, this approach helps to determine the total scale of the effects of these improvements by considering the total number of communities for which travel costs change as a result of the project.

VII. Outcomes Studied

MCC's investments in infrastructure improvements are aimed at generating higher economic activity and incomes. Projections of the impacts of road improvements are based on the aforementioned recent evidence from the evaluation literature identifying significant effects on household material well-being, with several of the studies pointing to gains in agricultural prices and production, lower input costs, and higher wages as the channels for income or consumption gains. MCC's impact evaluations will contribute to this growing body of evidence by analyzing a variety of outcomes from infrastructure improvements.

The outcomes studied in each of the impact evaluations discussed herein are related to the original economic analysis and to the purpose of the evaluation. As a standard across all countries, enough information is collected during and after implementation to calculate an *ex post* ERR using traditional methods, as discussed in Section II.

For roads investments, MCC collects the following common indicators across all countries:

- Average Annual Daily Traffic
- International Roughness Index
- Number of Kilometers Completed

These indicators allow for the calculation of changes in travel time and cost that result from the road improvements.

Beyond these measures of travel time and costs, the roads impact evaluations considered in this study assess the effects of these travel-specific changes on the broader economy. Because MCC's mission focuses on poverty reduction, many of these evaluations directly measure the change in economic well-being of individuals affected by the road improvements. As MCC's Chief Economist writes, "MCC's mandate starts with the recognition that income-based poverty measures are broadly accepted and used around the world; if an agency has as its mission the reduction of measured poverty, it needs to be able to demonstrate that it is raising incomes."[6] Because of this mandate, MCC's road impact evaluations are designed to estimate the change in household income resulting from a new or improved road.

Most of the household surveys implemented for this type of analysis are similar to the World Bank's Living Standards Measurement Study (LSMS)[7] and they include detailed modules collecting information on household consumption. Consumption is used more frequently than income since survey respondents are more likely to remember and feel comfortable sharing information on their consumption of goods rather than income. In addition, in rural areas, households have so many different sources of income that it is difficult to capture them all. Finally, consumption is preferred over income because it is smoother over time whereas income is more prone to change based on short-term factors. Analysis of intra-household allocation of resources is not conducted in a quantitative manner, but when practical, qualitative information is collected along with the impact evaluation survey data to understand how road improvements affect men and women differently.

Collecting data on household income or consumption can be challenging, time consuming and expensive, so in some countries the analysis is limited to changes in prices and availability of goods at local markets. The theory behind this is that improved roads will open up markets to more competition and the prices of goods will fall, which will automatically result in savings of income for households in that area. Often this strategy is used in countries that do not have an ongoing annual national household survey that would provide an easy basis from which to over sample for the evaluation. This approach also makes sense when the project includes improvements in rural roads for communities that before the project had very limited access to major markets. Another advantage to this approach is that changes in prices are expected to take place in the short-term whereas changes in income may take longer to develop.

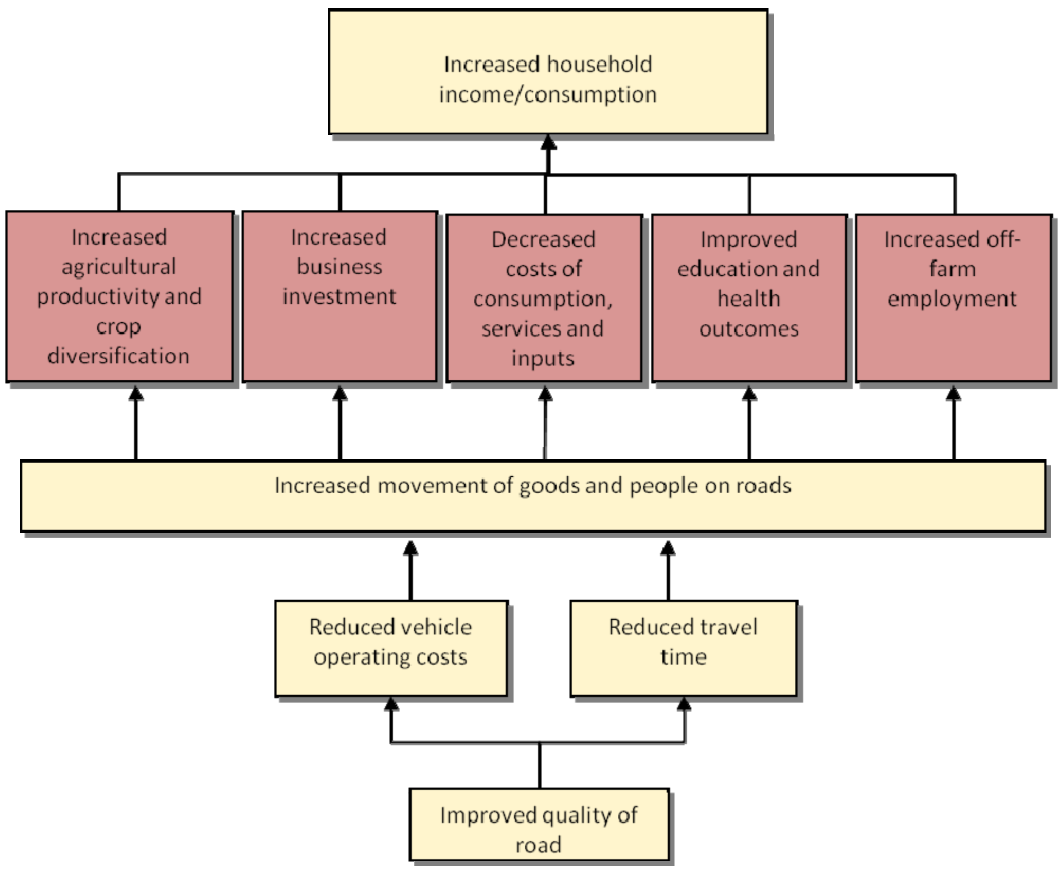
---

[6] Wiebe, Franck. MCC Working Paper "Aid Effectiveness: Putting Results at the Forefront" http://www.mcc.gov/mcc/bm.doc/mcc-112008-paper-results.pdf

[7] For more information visit www.worldbank.org/lsms

Certain evaluations are designed to answer not only the question of how household income/consumption has changed, but why. This is portrayed as the boxes shaded in red in the logic diagram below (Figure 3). For example, in El Salvador, the compact program includes upgrading and opening sections of the main road in the Northern Zone of the country, in a primarily agricultural area. The evaluation survey was designed to capture information not only on agricultural productivity and crop diversification, but also on employment and time use. The hypothesis is that as farmers gain greater access to nearby towns and cities, rural-urban trade increases, and markets become more integrated, opportunities for rural non-farm income generation should expand. It is expected that there will be changes in labor allocation between farm and non-farm activities.

Figure 3: Road Program Logic



Many countries project increases in business income from road improvements when they submit their project proposal justifications to MCC. Business income is not included in *ex ante* ERR estimates because of the lack of evidence and concern about double-counting; however, recognizing that it is an important area for more research, analysis of business development has been included in the road impact evaluations in Georgia and Honduras in addition to household income. Enterprise surveys are being conducted in addition to household surveys to capture business investment, employment, and income changes.

Likewise, the Tanzania and El Salvador evaluations include analysis of how the improved roads affect access to public services and their impacts on health and education outcomes. Table 5 summarizes the key outcomes included in each of the seven roads impact evaluations.

Table 5: Road Improvement Outcomes Studied

Country	Household Income / Consump.	Local Market Prices of Inputs/ Outputs	Business Income / Growth	Agric. Productivity	Transport access, time, costs	Access to health and education services	Land values	Additional Outcomes
Armenia	Y			Y	Y			
El Salvador	Y			Y	Y	Y	Y	Labor allocation between farm and non-farm activities
Georgia	Y		Y		Y			
Ghana		Y			Y			
Honduras	Y	Y	Y		Y	Y	Y	
Nicaragua		Y			Y			Availability of goods
Tanzania	Y	Y			Y	Y		Availability of transport services Capital investment Employment Migration patterns

An array of data collection instruments are being used to capture information on these outcomes, as illustrated in Table 6 below. All of the surveys involve panel data at either the community level or household level, a crucial element for the DD analysis described in Section III.

Table 6: Road Improvement Evaluation Surveys

Country	Data Sources	Data Collector	Survey Sample
Armenia	Integrated Living Conditions Survey	National Statistical Service of Armenia	Households selected by chance into main sample, plus an additional sample of 1,700 households served by project and comparison roads (panel of road sections)
El Salvador	Connectivity Household Survey and Community Survey	El Salvador Office of Statistics and Census	Panel of 5,388 households
Georgia	Integrated Household Survey and Village Infrastructure Census	Department of Statistics of Georgia	Over-sample of national household survey of 3,382 households (panel of communities)
Ghana	Ghana Market Survey	National Opinion Research Center and Pentax (private firms)	Panel of 308 communities
Honduras	Household Survey, Business Survey and Price and Product Survey	National Institute of Statistics of Honduras	2,000 households Panel of 200 enterprises Panel of 100 communities
Nicaragua	Price Survey	Fideg (private firm)	Panel of 435 observations in 33 communities
Tanzania	Household and community surveys	Economic Development Initiatives, Ltd (private firm)	3,000 households and panel of 200 communities for trunk roads 1,200 households for rural roads on Zanzibar

The road evaluations discussed herein will not only assess the average increase in income among broad populations from road improvements, but also will look at the heterogeneity in benefits throughout these populations. Understanding who receives these benefits is crucial for future policy decisions allocating scarce aid resources across potential projects. Traditional road analysis considers the users of the road to be the main beneficiaries and this is reflected in the *ex ante* economic analysis of time and cost savings for road users. However, for standard reporting across countries, MCC uses a common definition of road beneficiaries as the population that lives within 5 kilometers of the improved road. The reality is probably a mix of these two – road users benefit and people living near the road benefit as well. To understand who the main beneficiaries are of road improvements MCC is analyzing the impact of the road on households with different levels of access to the road. For example, the evaluation sample in Tanzania was purposefully stratified based on distance from the road, so that results could be compared between households within 30 minutes of walking time of the road and those farther than 30 minutes from the road. In addition, MCC conducts beneficiary analysis both pre- and post-construction, which disaggregates the increases in income from the ERR into income groups to estimate the distributional benefits of road improvements.

MCC's evaluations of investments in electricity, water and sanitation and irrigation will also involve analysis of income and consumption effects across the baseline distribution of beneficiaries. In monitoring the progress of these activities, MCC uses the following common indicators in all water and sanitation and irrigation projects when relevant (common indicators for electricity investments have not yet been developed):

Table 7: Common Indicators

Water and Sanitation	Irrigation
Persons Trained in Hygiene and Sanitary Best Practices	Hectares under improved or new irrigation
Number of Water Points Constructed	Hectares under production
Number of Sanitation Systems Constructed	
Volume of Water Produced	
Number of Households with access to Improved Water Supply	
Number of households with access to Improved Sanitation	
Domestic Water Consumption	
Commercial Water Consumption	
Incidence of Water-borne Diseases	

As discussed, MCC impact evaluations take the analysis a step further by analyzing the changes observed in households or businesses as a result of having, for example, increased water consumption and lower disease rates, and comparing them to a group who hasn't received the improvements. The household surveys, business surveys and community surveys used for these impact evaluations do not vary substantially from those used in MCC's road impact evaluations. In fact, in El Salvador, the same survey is being used to collect data for the roads evaluation and the rural electrification evaluation because of the overlapping locations of the two projects. Thus, the income and consumption effects of roads improvements can be compared to those generated from the electrification intervention. As noted previously, estimating the effects of water and sanitation programs on household income and consumption will be a major contribution of MCC's evaluations.

One difference between MCC's road and other infrastructure evaluations is that some of these other evaluations make use of administrative data such as utility company records. Information from the utilities on electricity used or water consumed and amount paid is collected to check household and business responses on surveys. Table 8 summarizes the key outcomes included in each of the non-roads infrastructure impact evaluations.

Table 8: Electricity, Irrigation, and Water and Sanitation Improvement Outcomes Studied

Country and Project	Household Income / Consump.	Business Income / Growth	Time / Cost Savings	Water / Electricity Consumption	Additional Outcomes
Armenia Irrigation	Y			Y	Agricultural productivity
El Salvador Electricity	Y		Y	Y	
El Salvador Water	Y		Y	Y	Health
Georgia	Y	Y	Y	Y	Health
Lesotho Water	Y	Y	Y	Y	Health
Tanzania Electricity	Y	Y		Y	Health, Education

MCC's evaluations of electricity, water, and sanitation investments will also address how time saved from daily household activities will be reallocated. There is generally consensus that providing an improved energy or water source will free up time for household members. However, whether this time will be allocated to leisure or to other productive activities remains less clear. Three of MCC's infrastructure evaluations discussed here have questionnaires that include time use modules to determine what effects time savings have on the household's productive activities.

Finally, even though a great deal of the existing literature on water and sanitation improvements in developing countries focuses on reductions in disease incidence, not all of MCC's evaluations highlighted herein are designed specifically to address this particular outcome. In Lesotho, the evaluation of the rural water supply program will indeed focus on gastrointestinal morbidity among children under 5 years of age. The evaluation of the urban water investments, however, will not, as urban water quality is already quite good and water-borne disease is not a significant problem. In El Salvador, MCC encountered a trade-off in terms of what outcomes could be analyzed with high precision. To capture the best data possible on coping costs (time costs associated with collecting water and the monetary cost of relying on alternative water sources such as vendors), surveys would have had to be done in the dry season when water shortages are more likely. To capture data on disease incidence, however, surveys would have had to be done during the wet season when disease rates are highest. Given budget limitations, MCC could not commission dual surveys in both seasons and opted to give priority to the measurement of changes in household income and coping costs by conducting surveys in the dry season. This was due in large part to the fact that coping costs accounted for most of the expected benefits in the ERR and because diarrhea rates had dropped substantially in recent years due to a variety of public health interventions.

VIII. Implementation

Even though none of MCC's infrastructure impact evaluations have been completed yet, there are many lessons from the past 6 years about how implementation can affect impact evaluation design – ranging from creating opportunities for evaluation where there was none to preventing the opportunity for any rigorous evaluation at all. The impact evaluation concept is developed before a compact is even signed. Usually the final evaluation design is agreed upon during the first year of implementation once implementation strategies are better defined. The final agreed upon design is completed in the first year of the compact term so that baseline data can be collected before construction begins. Implementation plans change throughout all 5 years of the compact term. So far, three common issues have been encountered during implementation that complicate the impact evaluation effort. Those are: (1) project scope changes because of cost over-runs and/or poor policy performance; (2) delays in construction; and (3) changes to the roll-out and/or contracting strategy. Each one is discussed in turn.

In many countries, compacts are signed before feasibility studies, detailed designs, environmental impact assessments, and resettlement plans are completed. Therefore, at compact signing when the impact evaluation design is initially agreed upon, and the scope of infrastructure to be improved (the number of kilometers of roads, for example) is usually an upper bound. As studies are completed, the number of kilometers to be completed is refined. This takes place over a couple of years. Between collecting baseline data and finalizing the detailed design of the roads, currencies fluctuate and input prices change. In the recent past, input prices have increased, so the reasonable pre-investment cost estimates have turned into underfunded budgets and fewer road sections have been improved. Given a different global financial situation, one could imagine that the opposite could take place – more road sections than originally planned could be constructed.

In addition to designs, currency fluctuations, and input price changes, projects can be changed because of poor policy performance by a country. As explained in Section II, eligibility for MCC assistance is based on countries meeting certain requirements in terms of ruling justly, investing in people, and economic freedom. If countries backslide on policy performance during implementation, MCC funding can be put on hold, suspended, or terminated. This occurred in two of the countries discussed in this paper – Armenia and Nicaragua.

The effect of project scope changes or "re-scoping" on the impact evaluation could be either positive or negative. As road projects get smaller in terms of the number of kilometers completed and the number of beneficiaries, the treatment group for the evaluation also gets smaller. If the baseline survey was completed and then a re-scoping takes place and the treatment survey sample is reduced, the power of the evaluation is reduced. As explained in Section IV, in Armenia, an RD design was chosen because road sections had been packaged for construction in three packages based on ERRs. However, subsequent challenges in the country led MCC to put a hold on the road activities. Only 24 kilometers were built by MCC, which was too small of a treatment group to conduct the originally planned impact evaluation.

In at least one case thus far, re-scoping has led evaluators to consider including in the comparison group road segments that would no longer be improved. In Nicaragua, sections that were initially slated for improvement but were not launched when compact funding was terminated were relatively comparable to those roads on which construction had already begun. Nonetheless, in order to make use of such evaluation opportunities, evaluators must carefully investigate whether the differential timing of improvements along particular segments is correlated with factors related to outcomes along these roads.

Preparing for road construction requires a large effort, especially when done under the MCC model of country ownership. Countries are required to develop their own terms of reference for construction and construction supervision, as well as all of the environmental assessments and resettlement plans. MCC then reviews and approves each of these. This process often takes months and has resulted in projects starting later than originally planned in some cases. Delays can also be caused by poor performance by design contractors, new information about environmental considerations, resettlement negotiations, or problems encountered during construction. All of MCC's road projects must be completed by the end of the compact, which has a 5 year term. It is very common for construction projects to be delayed such that the construction is completed in the last year of the compact.

Delays create a challenge for evaluating the impact of the road improvements. An evaluation at the end of the compact will only capture short-term effects and if the project is in an agricultural area, at least a year will have to pass for any agricultural changes to take place before doing the final data collection. In addition, if an evaluation based on differential timing of construction was planned, construction delays could end up condensing construction into a short time period towards the end of the compact term, making this type of analysis impossible. There are also risks for matching evaluations. If a matching evaluation was planned and the treatment takes years to be completed, it is possible that other road projects funded by the government or other donors could be started and completed in the comparison group at the same time as the MCC treatment group.

A third complication is changes to the project roll-out and/or contracting strategy. An evaluation based on differential timing of construction requires that certain road segments are improved before others. Sometimes final designs including environmental issues and resettlement become more complicated on a certain section than originally thought. This could delay construction on a particular section and completely change the order of section completion. If the construction coincides with later sections, the analysis based on differential timing of construction can no longer be implemented. Changes to contracting strategies can also result in schedule changes if road sections were packaged differently than originally planned. Combining construction packages could complicate an evaluation based on the differential timing of construction, since it would reduce the statistical power. Conversely, separating construction packages into smaller units could create an opportunity for analysis based on differential timing of construction (although timing between them may not be sufficient).

The evaluation issues experienced in implementation of road improvement projects have been similar to those experienced by projects focused on other types of infrastructure. In El Salvador, both the water and sanitation and rural electrification evaluations were designed based on pre-feasibility studies and as final designs were completed, scope changes occurred, which caused several complications for the evaluations. In water and sanitation, many of the treatment projects were determined to be unfeasible and dropped from the project, leaving the evaluation with a smaller treatment sample than expected. As a result, a second baseline including the final list of treatment projects may have to be undertaken, as the first baseline only included a sample of expected projects (luckily for the evaluation, implementation is behind so there is time for another baseline). In the case of El Salvador's rural electrification evaluation, some of the final designs for extending the electrical grid left out households included in the baseline survey as part of the treatment group based on the pre-feasibility designs. The evaluation team is working with the implementation team to identify these households and include them in future phases of electrical grid extensions when possible.

IX. Recommendations

The 13 MCC evaluations studied herein offer useful lessons for evaluators and are likely to generate several key pieces of information that will guide policymakers considering investing resources in potential infrastructure improvement projects. Most notably, the evaluations in El Salvador, Honduras, Nicaragua, Georgia, and Tanzania are likely to provide the first set of empirical evidence on the impacts of improvements in highways and major secondary roads in developing countries. Evaluations of rural road improvements will help quantify the effects of these improvements on the access to and use of public services, such as medical clinics, as well as on other channels through which households' material well-being may be affected. MCC's evaluations of water and sanitation improvements will include some of the first experimental and quasi-experimental results on the effects on current income generating opportunities for households in developing countries. Meanwhile, electrification evaluations will provide rigorous evidence on the extent to which access to electricity leads to changes in the level and composition of household income, including the wages earned by women and the amount of unpaid work they do at home.

The MCC portfolio of evaluations also offers several more immediate implications for infrastructure evaluation. First, one important question not tackled by these evaluations is what effect MCC's engagement with recipient countries on reforming institutional arrangements for maintaining the improved infrastructure has had on the affected population. For example, new road maintenance policies are being implemented along with MCC's capital investments in a number of countries. Although such issues are frequently national in scope and variation in treatment is often difficult to identify, one might think that these reforms would differentially affect areas where existing roads had languished longest without adequate maintenance. Such heterogeneous effects could help evaluators assess the likely value of such reforms.

The second implication is that comparison of non-experimental and experimental methods for evaluation rural road improvements is crucial to confirming the validity of the former. The overlap of methods within a number of the MCC evaluations will allow for a comparison of these methods. In particular, the El Salvador roads evaluation's RD design could be compared to estimates obtained through PSM-DD, giving a sense of the effectiveness of PSM-DD in providing precise, unbiased estimates of program impact. At the same time, it would be invaluable to compare these alternative approaches to experimental results. While random assignment is being used in several evaluations of water and electricity network expansion, future evaluators—particularly those at MCC—should continue to consider random assignment of rural road segments as a potential evaluation method. Random assignment of rural road segments may be most appropriate in cases where segment-by-segment cost-benefit analysis is not possible. In such a case, a comparison of the random assignment results to those obtained from PSM-DD and continuous treatment estimators would be extremely valuable to the roads evaluation literature.

Finally, MCC's evaluation experience suggests that evaluators must build multiple methods into their evaluation designs if they are to remain robust to project implementation changes. Evaluations that rely on selection processes or construction timing can offer some of the most rigorous results, but are naturally sensitive to changes in these processes or timing. On the other hand, evaluations using statistical matching are much more robust to such changes, though the degree of rigor they offer depends on how much of the treatment-comparison differences one can directly observe. Combining these methods in evaluations is thus highly recommended.

References

Behrman, J., Y. Chen and P. Todd (2004). "Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach" *Review of Economics and Statistics* 86: 108-132.

Chaplin, D., M. Chatterji, and D. Hankinson (2009). "Final Design Report for the MCC Impact Evaluation Design & Implementation Services – Tanzania Energy Sector Project."

Dehejia, R. and S. Wahba (1999). "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94: 1053-1062.

Dercon, Stefan, Daniel Gilligan, John Hoddinott, and Tassew Woldehanna (2006). "The Impact of Roads and Agricultural Extension on Consumption Growth and Poverty in Fifteen Ethiopian Villages." CSAE WPS 2007-01, University of Oxford, UK.

Economic Development Initiatives Ltd (2009). "Final Report: Consultancy Services for the Design and Implementation of Household Survey and Community Profile for Transport Sector."

Escobal, Javier and Carmen Ponce (2004). "The Benefits of Rural Roads: Enhancing Income Opportunities for the Rural Poor." GRADE Working Paper 40, Lima Peru.

Estache, A (2010). "A Survey of Impact Evaluations of Infrastructure Projects, Programs and Policies." ECARES Working Paper 2010-005, Université Libre de Bruxelles.

Fideg (2008). "Informe Final: Encuesta para medir el precio de la Canasta Básica en 30 Comunidades de León y Chinandega."

Fortson, Ken and Anu Rangarajan (2008). "Rural Road Rehabilitation Project Evaluation Plan." Mathematica Policy Research.

Heckman, J., H. Ichimura, and P. Todd (1998a). "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies* 65: 261-294.

Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998b). "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66(5): 1017-1098.

Hirano, K., and G. Imbens (2004) "The Propensity Score with Continuous Treatments" in A. Gelman and X. Meng (eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, Wiley.

Imbens, G. (2000) "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika* 87: 706-710.

Imbens, G. and J. Wooldridge (2008). "Recent Developments in the Econometrics of Program Evaluation." National Bureau of Economic Research Working Paper 14251.

Jalan, S. and M. Ravallion (2003). "Does Piped Water Reduce Diarrhea for Children in Rural India." *Journal of Econometrics*, 112(1): 153-173.

Khandker, Shahidur, Zaid Bakht and Gayatri Koolwal (2006). "The Poverty Impact of Rural Roads: Evidence from Bangladesh." World Bank Policy Research Working Paper 3875.

Komives, K., C. Poulos and S. Pattanayak (2009). "Impact Evaluation Design for MCC Water and Sanitation Interventions in El Salvador." Social Impact, Inc.

Lee, L., M. Rosenzweig, and M. Pitt (1997). "The Effects of Improved Nutrition, Sanitation, and Water Quality on Child Health in High-Mortality Populations." *Journal of Econometrics*. 77(1): 209-235.

Lokshin, Michael and Ruslan Yemtsov (2005). "Has Rural Infrastructure Rehabilitation In Georgia Helped the Poor?" The World Bank Economic Review 19(2): 311-333.

National Opinion Research Center (2007). "Design Report: Design and Implementation of MCA Honduras Program Evaluation."

National Opinion Research Center (2009a). "Annual Report: Impact Evaluation Design & Implementation Services Georgia."

National Opinion Research Center (2009b). "Department of Rural Water Supply Evaluation Mini-Report (Impact Evaluation Design & Implementation Services – Lesotho)."

National Opinion Research Center (2009c). “Water and Sewerage Authority Evaluation Mini-Report (Impact Evaluation Design & Implementation Services – Lesotho).”

Rauniyar, G., A. Orbeta and G. Sugiyarto (2010). “Impact of Water Supply and Sanitation Assistance on Human Welfare in Rural Pakistan.” Asian Development Bank.

Rosenbaum, P. and D. Rubin (1983). “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70(1): 41-55.

Struyk, Raymond, Joseph Caldwell, John Felkner, Kareem Kysia, and Shova KC (2010). “Impact Evaluation of Feeder Roads: Phase 1, Baseline Findings Report.” National Opinion Research Center.

Torero, Maximo (2009a). “Impact Evaluation Design for MCC Connectivity Interventions in El Salvador.” Social Impact, Inc.

Torero, Maximo (2009b). “Impact Evaluation Design for MCC Rural Electrifications Interventions in El Salvador.” Social Impact, Inc.

Van de Walle, D. & D. Cratty (2002). “Impact Evaluation of a Rural Road Rehabilitation Project.” World Bank, Mimeo.

Van de Walle, D. (2008) “Impact Evaluations of Rural Road Projects.” World Bank Working Paper 47438.

White, Howard (2008), ed. *An Impact Evaluation of India's Second and Third Andhra Pradesh Irrigation Projects: A Case of Poverty Reduction with Low Economic Returns*, World Bank: Washington, DC.

Wiebe, Franck (2008). “Aid Effectiveness: Putting Results at the Forefront.” MCC Working Paper.