MATHEMATICA
Policy Research, Inc.

MEMORANDUM

P.O. Box 2393
Princeton, NJ 08543-2393
Telephone (609) 799-3535
Fax (609) 799-0005
www.mathematica-mpr.com

TO: Rebecca Tunstall

FROM: Ken Fortson; Anu Rangarajan

DATE: 6/15/2010
MCC-Armenia

SUBJECT: Rural Road Rehabilitation Evaluation Plan

The Millennium Challenge Account with Armenia (MCA) aims to increase household income and reduce poverty in rural Armenia through improved performance of the country's agricultural sector. MCA planned to achieve this goal through an integrated, nationwide initiative to improve major components of rural infrastructure, focusing on rural roads and the irrigation system, supplementing these infrastructure improvements with training in agricultural practices. By improving living standards among rural residents, these investments could in turn lead to economic growth in rural areas and throughout the country as a whole. The Millennium Challenge Corporation (MCC) has commissioned a rigorous impact evaluation to separately examine each of the three main components (roads, irrigation, and agricultural training/support) of its compact with Armenia (the "Compact").

This memo describes our updated design for evaluating the impact of roads that were planned to be included in MCA's Rural Road Rehabilitation Project (RRRP). The original evaluation design planned to compare roads selected for rehabilitation to road projects that were considered but not selected for rehabilitation by MCA. The evaluation design would have leveraged the systematic process MCA used to select roads for rehabilitation to estimate credible program impacts.

However, MCC funding for the RRRP was put on hold indefinitely due to concerns about the Armenian government's commitment to democratic governance, stemming from the government's crackdown on protests following the last presidential election. As part of its own rehabilitation effort, the World Bank decided to fund the rehabilitation of some but not all of the road links included in MCA's plans. For the roads it funded, the World Bank followed the road project designs developed by MCA, with some of the designs updated due to changes in conditions since designs were developed (about two years ago). The revised impact evaluation will estimate impacts by focusing primarily on road projects originally designed by MCA but now funded by the World Bank.

In the sections that follow, we discuss our proposed research design in more detail. The key research questions guiding our design of the RRRP evaluation are:

- Did rehabilitating roads affect the quality of roads?

- Did rehabilitating roads improve access to markets and social services?
- Did rehabilitating roads affect agricultural productivity and profits, and if so, by how much?
- Did rehabilitating roads improve household well-being for communities served by these roads, especially income and poverty?

We start by describing our original evaluation design in Section I, followed by a discussion of the comparison group design that serves as the foundation of our revised analytic approach in Section II. We then discuss the main source of data for the roads evaluation, the Integrated Living Conditions Survey (ILCS) in Section III. We next discuss in detail our econometric approach for estimating road rehabilitation impacts in Section IV. Section V provides an approximate timeline for the impact evaluation report. Finally, we outline some information we would like to obtain to help finalize the evaluation design in Section VI.

I. ORIGINAL EVALUATION DESIGN

A crucial objective of any impact evaluation is to assess not only how key outcomes for affected beneficiaries changed after the intervention, but also to assess the counterfactual: how outcomes *would have changed* even in the absence of the intervention. Randomly assigning the intervention is the ideal way to identify a credible counterfactual. For instance, randomly selecting which roads would be rehabilitated from amongst the eligible is an ideal evaluation design because the roads that are not selected under random assignment should not differ, on average, from those that were selected and, thus, serve as a compelling counterfactual.

However, as is frequently the case for large scale infrastructure projects, it was not feasible to randomly select the roads that would be rehabilitated in the MCA RRRP from amongst the eligible. Instead, MCA planned to fund projects based on estimated economic rates of return (ERRs). The ERR is calculated from several inputs, including the vehicular traffic, vehicle operating costs, and the cost of the project, among others. An ERR was estimated for each of the road links under consideration for rehabilitation, and in order to be funded, a road's ERR had to meet or exceed 12.5 percent.

Though it precluded random assignment, the systematic selection of roads for rehabilitation introduced an opportunity for another rigorous evaluation design. In particular, this arrangement lent itself to a *regression discontinuity* approach that would exploit the economic rate of return (ERR) threshold that MCA used to determine which road link would be funded. In a regression discontinuity design, treatment is systematically assigned based on a known "score", which in this context is the ERR; impacts are measured by comparing outcomes for those above and below the score selection threshold. Because the exact selection criteria are known and can be accounted for in the statistical analysis, regression discontinuity is considered a strong, credible alternative when random assignment is not possible. In this context, the research design would

have essentially compared roads that just passed the ERR threshold to those that fell just short of it. The roads on either side of this cut-off have very similar ERRs, yet roads above the threshold would be rehabilitated, and roads below would not be. For roads near the threshold, *ex ante* differences between the two groups are close to random, and therefore outcome differences for roads on either side of the thresholds would reflect program impacts.

The data for this evaluation would have come from Armenia's Integrated Living Conditions Survey (ILCS), an annual household survey fielded by Armenia's National Statistical Service (NSS). MCA is funding a considerable increase in the rural sample, which was in part aimed at facilitating the RRRP evaluation. The sampling plans were developed and implemented based on the evaluation design described above. By design, the survey sample includes communities from 82 of the 85 road links MCA considered for rehabilitation, and estimates suggested that about half of these would meet the ERR threshold for rehabilitation. We discuss the content and design of the ILCS in greater detail in Section III.

II. COMPARISON GROUP DESIGN

After the MCA road rehabilitation program was put on indefinite hold, the World Bank decided to fund rehabilitation of some of the road links that had been included in the RRRP plans. (The Bank is also concurrently funding other road rehabilitation projects that were not considered under the RRRP.) The World Bank is basing the rehabilitation efforts on project designs developed by MCA, with some of the designs updated due to changes in conditions after the designs were developed about two years ago. Implementation ofc these projects began in late 2009 and will continue through 2010. Our revised evaluation design focuses on road links that were formerly in the RRRP and are now funded by the World Bank. The evaluation will also include a couple of roads that were rehabilitated through MCA before the program was put on hold. Like MCA, the World Bank requires that the projects it funds be economically justified. However, in selecting which of MCA's original project designs would be funded by the World Bank, the used other criteria in addition to economic rates of return.

Consequently, the regression discontinuity evaluation design we developed is no longer a viable approach to measuring the impact of road rehabilitation. Instead, the evaluation of rural road rehabilitation will use a comparison group design. The revised evaluation design identifies the counterfactual by defining a comparison group of roads selected from those initially proposed by MCA that share similar characteristics as those selected by the World Bank, but that were not rehabilitated.

In a comparison group design, a non-randomly selected set of comparison communities is employed to estimate the counterfactual. In this case, impacts will be estimated by comparing treatment group communities, where roads were rehabilitated, to a comparison group of communities, where roads were not rehabilitated. An important concern with a comparison

group design is that, because the treatment and comparison communities are not randomly selected, there may be observed or unobserved differences on some dimension that affect the key outcomes. More specifically, the reasons a given road was selected for rehabilitation (such as having a large population of beneficiaries) may also contribute to differences in average outcomes for the two sets of communities. To the extent that the comparison group differs from the treatment group along dimensions important for outcomes, we will be unable to distinguish actual program impacts from underlying differences in the two groups unless we can credibly identify and control for these pre-existing differences.

There are two common methodologies to account for differences between the treatment and comparison groups: statistical matching and regression-based methods. Both of these methods adjust for differences in observable characteristics between the two groups, but they each have advantages and limitations. Statistical matching is usually the preferred approach theoretically because it can create a set of comparison communities with very similar observable characteristics, on average, to the treatment communities. Though regression-based methods can also control for differences between the groups, they are dependent on correctly modeling underlying relationships between the outcome measure and baseline characteristics, whereas statistical matching obviates this need when the treatment and comparison group have very similar sets of characteristics. However, as a practical matter, statistical matching can be challenging, since close matches for each treatment community can only be identified if there is a large pool of possible comparison communities from which to draw. Because we have a limited pool of comparison communities, we will adopt a hybrid approach that matches road links based on a limited number of key observable characteristics and then uses regression modeling to control for other characteristics that may impact outcomes.

The most fundamental criterion we will use in identifying potential comparison roads is that it is among the roads that were originally considered for rehabilitation by MCA, indicating that the road conditions are poor enough to consider rehabilitation. Furthermore, MCA's original list of eligible roads was divided into construction packages based largely on traffic counts, and nearly all of the World Bank-funded roads come from one of these construction packages (specifically, the second). Consequently, we will further limit the comparison group to the unfunded road links within the same original MCA package. Matched comparisons will be selected from among the unfunded roads in the same construction package, matching on geographic region where possible. In total, we project that we will be able to include about 20 treatment and 20 comparison roads in the analysis. The precise number will be determined after consulting with MCA about exactly which road links can be included in the analysis, as discussed in Section VI.

III. INTEGRATED LIVING CONDITIONS SURVEY

The key outcome data for the RRRP impact evaluation will come from the ILCS. The ILCS is an annual, nationally representative, repeated cross-sectional household survey fielded by NSS and covering a broad range of topical domains including demographics, employment, income, agriculture, and public services, among others. The core sample of the ILCS includes 768 enumeration areas, each containing 8 households for a total sample of approximately 6,100 households. Additionally, MCA is funding an oversample of 216 enumeration areas in rural communities, an increase of approximately 1,700 households.

The oversample is dedicated exclusively to communities served by roads in MCA's initial set of rehabilitation-eligible roads.[1] Additional communities served by rehabilitation-eligible roads were selected into the core sample by chance, and a total of approximately 2,200 households served by the original, eligible project roads are in the sample each year. The sample includes communities served by 82 of the 85 original eligible roads. (The three other project roads were excluded to preserve geographic balance in the ILCS oversample.) Because the ILCS does not use the same sample of enumeration areas in each year of the survey, different rounds of the survey may include different communities served by these roads. Based on our projections discussed in Section II, we expect that the revised evaluation design will focus on communities served by 40 eligible roads (20 treatment, 20 comparison), yielding a sample of about 1,075 households.

The standard ILCS questionnaire includes questions on several key outcomes relevant to the evaluation, including household income. This questionnaire was expanded during the evaluation period to include questions about intermediate outcomes related to the MCA programs—including the RRRP—as well as an expanded set of final outcomes. For the RRRP evaluation, the ILCS provides data on numerous intermediate and final outcomes, as described below.

Intermediate Outcomes. While most of the outcomes of primary interest to MCA and MCC are longer-term outcomes, such as economic improvements (including household income), these outcomes may not be immediately observable. Consequently, we will closely examine intermediate outcomes through which the road rehabilitation projects are intended to improve household income; these outcomes may be more likely to change over the (limited) study period.

---

[1] The ILCS sampling approach only includes communities *directly* served by project-eligible roads; however, other nearby communities may benefit from rehabilitation as well. Because impacts are likely largest in communities directly served by project roads, the impact evaluation is geared toward maximizing statistical precision for those communities. Hence, the estimated impacts will be interpreted as the average impact on households in communities that are directly served by project roads.

We would expect an impact on households' income only if we observe that a substantial proportion of the targeted communities are actually experiencing improvements in their roads. Examining intermediate outcomes also establishes the counterfactual—how the quality of and access to roads would have changed even in the absence of the road projects. Table 1 summarizes the key intermediate outcomes that can be examined using the ILCS data.

Table 1. Intermediate Outcome Measures

Intermediate Outcome Measures	Time Frame
Utilization of Local Roads. Whether (and how often) road transportation is typically used to purchase agricultural supplies, to sell agricultural produce, to access employment outside the community, or for other purposes	Typical Month
Perceived Quality of Roads. Quality of roads both within the community and roads connecting the community to other communities; quality of local public transportation.	As of Survey Date
Availability of Transportation. Modes of transportation are available and commonly used, especially to access key community services.	As of Survey Date
Access to Social Infrastructure. Distance and time from key services, including health facilities, schools, community centers, and markets.	As of Survey Date
Agricultural Challenges. Whether transportation or road quality impeded access to markets; whether potential produce buyers visited communities more often after rehabilitation.	Last Agricultural Season

Final Outcomes. The ultimate goal of the RRRP is to increase household income in rural Armenia, and hence, household income and its components are an important focus of our analysis. The primary domains we will focus on are agricultural revenue, costs, and profits and employment income by household members, as improvements in these outcomes were central to MCA's conceptual framework for the Compact. We can also use the average sale price of specific crops for other farmers in the community to monetize crops that are produced by the household, but then consumed or bartered. Rehabilitating roads might also increase non-agricultural income, especially employment opportunities; therefore, we will also estimate impacts on key sources of non-agricultural income. Table 2 summarizes the key final outcomes that can be examined using the ILCS data.

Table 2. Final Outcome Measures

Final Outcome Measures	Time Frame
Agricultural Production. Total amount of specific crops grown; amount of crops grown per square meter; total value of all crops cultivated; total amount consumed.	Last Agricultural Season
Food Production. Total amount of specific foods produced, such as bread, cheese, and milk; total value of all food produced; total amount consumed.	Last Agricultural Season
Livestock. Number of cows, pigs, sheep, and other livestock owned as of the survey date; number of livestock bought and sold in last 12 months.	As of Survey Date / Last Year
Revenue from Agricultural or Food Production. Value of crops or food sold; total value of all crops or food (including those sold, bartered, or consumed).	Last Agricultural Season
Agricultural Costs. Expenditures on fertilizers, pesticides, irrigation water, hired labor, equipment, and taxes (individually and in total).	Last Agricultural Season
Profit from Agricultural and Food Production. Revenues minus costs—the income from agricultural activities and food production. This could be measured two ways: One would be the purely monetary profit, while the other would monetize the amount consumed by the household, including this is profits.	Last Agricultural Season
Income from Employment. Whether household head, spouse, and any grown children were employed; total earnings from employment.	Last Week
Income from Remittances.	Last Year
Household Income. Sum of agricultural profits, employment income, and remittances; poverty rate.	Last Year

IV. ESTIMATING PROGRAM IMPACTS

In this section we describe the empirical models that will be used to estimate impacts of the implemented road rehabilitation projects. The ILCS survey data is cross-sectional, with a new cross-section of respondents (and enumeration areas) drawn each year. Given this data structure, the specific empirical framework we will employ is commonly known as *difference-in-differences* estimation. The central idea behind difference-in-differences estimation is to estimate how outcomes changed for communities served by treatment roads before and after road rehabilitation (the first difference) and compare this to how outcomes changed for communities served by comparison roads over the same time period (the second difference). The difference between these two differences is the estimated impact that can be attributed to the program. Baseline (pre-intervention) data will come from the 2007 and 2008 rounds of the ILCS; follow-

up (post-intervention) data will come from the 2011 round of the ILCS.[2,3] Impacts can be estimated mathematically using the following regression model:

$$(1)\ y_{irt} = \beta' x_{ir} + \lambda T_r + \theta' Y_t + \gamma T_r \times Y_{t\neq 2011} + \eta_{rt} + \varepsilon_{irt}$$

where y_{irt} is the outcome of interest for household i served by road r at time t; x_{ir} is a vector of time-invariant characteristics of household i served by road r; T_r is an indicator equal to one if road r is in the treatment group and zero if it is in the comparison group; η_{rt} is a road-specific error term (a road "random effect"); ε_{irt} is a random error term for household i served by road r observed at time t; and β, ϕ, γ, λ, δ, and θ are parameters to be estimated.

The vector of baseline characteristics x_{ir} will include both household and community characteristics. At a minimum, we will control for community characteristics such as the geographic region or the ERR originally estimated for its road. We will also control for household size and composition, and characteristics of the household head, namely, education level, gender, and age. In the framework of a repeated cross-sectional model, however, the characteristics that are included must be restricted to those that are unaffected by the road rehabilitation project.

The ILCS was fielded in two years prior to program implementation (2007 and 2008), and we can use both years of pre-intervention data on key outcomes—rather than just one year—to improve precision of the impact estimates.. The vector Y_t contains indicator variables for each survey year, accounting for countrywide trends that affect all households in a particular year. The coefficient γ is an estimate of how different outcomes were in the treatment and comparison communities *prior to rehabilitation*. In this formulation, the estimate of the parameter λ is the difference-in-differences impact estimate, and represents the difference in outcomes for treatment and comparison communities in 2011, the first year after projects are complete in all of the beneficiary communities, relative to how different they were prior to rehabilitation. Standard t-tests can be used to gauge the statistical significance of the impact estimates.

Clustering. Outcomes for households served by the same road (and in the same community) cannot be considered statistically independent, as neighboring households may be exposed to

---

[2] MCA funding for the ILCS is currently scheduled to expire in the fall of 2011. NSS, MCA, and MCC are exploring whether the current ILCS staffing and sample sizes could be sustained through the end of 2011.

[3] Because the programs were implemented in 2009 and 2010, the ILCS for these two years will not be included in the analysis.

similar idiosyncratic influences that are not otherwise captured in the regression model. As an example, a particular community might have abnormally good or bad weather, or could experience other economic "shocks" that are unrelated to the road project but nonetheless affect the entire community. The econometric models will account for this clustering with methods that allow flexibility in the correlation structure of the error terms.

Precision of the Impact Estimates. A limitation of the impact analysis is that our estimates will not be very statistically precise because of features of the evaluation design. Relative to an unclustered sample design, the clustered nature of the intervention reduces statistical power. Instead of having 1,075 households drawn independently from many different communities, they are drawn from just 40 road projects. We estimate that the minimum detectable impact on poverty under this evaluation design is 11 percentage points.[4] The limited statistical power is especially important because we will be estimating impacts based on data collected one year after many of the programs were completed. Limited statistical power means that we will not be able to detect small impacts, as we might expect at the one-year follow-up (when full impacts have not been realized). We will emphasize this limitation when we present the estimated impacts, and as discussed in Section III, it will be important to examine intermediate outcomes to see if there are indications that larger impacts on household wellbeing may develop in the future.

Tables. The impact report will include a variety of tables and figures with descriptive statistics on the data; however, the focus will be on estimates of road rehabilitation impacts and their statistical significance. We will also report regression-adjusted mean outcomes for the treatment and comparison communities—that is, the means for the two groups if they had identical community and household characteristics. Table 3 provides an example of the structure of these tables.

Table 3. Example of Impact Estimates Table

	Treatment Mean	*Comparison Mean*	*Program Impact*	*p-Value of Impact*
Household Income				
Poverty Rate				
Etc.				

Notes: */**/*** indicate statistical significance at the .10/.05/.01 level.

[4] These calculations assume a two-tailed t-test with 80 percent power and a 5 percent significance level; 20 treatment roads (and, therefore, clusters), 20 comparison roads (and, therefore, clusters), and 1075 total households responding; and an intraclass correlation of 0.062.

Impacts on Subgroups. For many of the outcome measures, it is conceivable that the effects of the road projects will vary by observable characteristics. Estimating differential impacts on female-headed households, for example, is of particular interest to MCC. We will examine whether the interventions' effects differ for key subgroups defined by household characteristics such as gender, age, and level of education of the household head; size of the household; or size of farm holdings operated by the household.[5]

It is straightforward to embed subgroup estimates into the framework of our previous specification. To do so, we include an interaction term that distinguishes treatment group members in subgroup *S* from those who are not in the subgroup:

$$(2)\quad y_{irt} = \beta' x_{ir} + \lambda_{S=1} T_r \times (S_{ir} = 1) + \lambda_{S=0} T_r \times (S_{ir} = 0) + \theta' Y_t + \gamma T_r \times Y_{t \neq 2011} + \eta_{rt} + \varepsilon_{irt}$$

In the equation above, the estimate of $\lambda_{S=1}$ represents the estimated impact for members of subgroup *S*, and we can test whether the impacts differ for members of that subgroup compared to households not in that subgroup by statistically testing whether $\lambda_{S=1}$ and $\lambda_{S=0}$ are equal.

V. PLANS FOR REPORTING

The final impact evaluation is planned for 2012, based primarily on data from the 2011 ILCS. The projected timeline for the evaluation is as follows:

2007-2008: Baseline ILCS data collection.
Late 2009-2010: Construction
2011: Final ILCS data collection
2012 Q2: NSS data processing completed
2012 Q4: Impact evaluation report completed

[5] In the original evaluation design, we also suggested subgroup analysis for subgroups defined by road link characteristics (such as region), not just by household. However, because of the smaller number of road links that will be rehabilitated with World Bank funding, it is not feasible to estimate impacts separately for subgroups of road links.

VI. INFORMATION REQUESTS

The evaluation design presented in this memo was developed based on our understanding of the progress of the road rehabilitation projects and the implementation plans for the remaining road links. To finalize this evaluation design, we kindly ask for MCA's assistance on the following pieces of information:

- Confirm that the roads we identified as World Bank-funded roads from the original MCA roads are correct, and check on a few related uncertainties;
- On the same list, identify the roads that were rehabilitated by MCA (and when) before the RRRP was suspended; and
- Match the roads to the communities included in the ILCS for each year.

We also welcome any corrections on details that we have misunderstood or omitted from this memo that would have implications for the evaluation design.