MATHEMATICA Policy Research

REPORT

DRAFT REPORT

# Evaluation of the Irrigation Infrastructure Activity in Armenia: Design Report

October 6, 2014

Kenneth Fortson
Joanne Lee
Madeline Young

Submitted to:
Millennium Challenge Corporation
875 Fifteenth St., NW
Washington, DC 20005-1203
Project Officer: Rebecca Goldsmith

Submitted by:
Mathematica Policy Research
505 14th Street, Suite 800
Oakland, CA 94612-1475
Telephone: (510) 830-3700
Facsimile: (510) 830-3701

Project Director: Kenneth Fortson
Reference Number: 06916.083

## CONTENTS

## TABLES



## FIGURES

## LIST OF ACRONYMS

| | |
|---|---|
| AREG | Armenian non-governmental organization that, in partnership with Jen Consult, is responsible for household-level data collection |
| AVAG | Armenian company that is responsible for collection of administrative data from water user associations |
| ERR | Economic Rate of Return |
| HVA | High-Value Agriculture |
| IPW | Inverse Probability Weighting (based on estimated propensity scores in the present context) |
| ISSA | Institutional Strengthening Subactivity |
| MCC | Millennium Challenge Corporation |
| MCA | Millennium Challenge Account (for Armenia) |
| OFWM | On-Farm Water Management |
| WtM | Water-to-Market |
| WSA | Water Supply Agency |
| WUA | Water User Association |

# I. OVERVIEW OF THE IRRIGATION INFRASTRUCTURE PROJECT

The Republic of Armenia was left with the legacy of a centrally planned economy when it declared independence from the Soviet Union in 1991. The Armenian economy was highly dependent on its Soviet trading partners and poorly equipped to function with the lack of infrastructure investment and support after Soviet withdrawal. In 1994, the Armenian government adopted a comprehensive stabilization and reform program that dramatically lowered inflation and led to steady economic growth beginning in 1995. Evidence from the Integrated Living Conditions Survey, however, suggests that this growth occurred primarily in urban areas. As of 2004, the poverty rate in rural areas was 32 percent (National Statistical Service, 2010).[1]

Although many rural Armenian households are involved in farming, farmers cannot rely on timely and adequate water to cultivate their crops. Much of the irrigation infrastructure established prior to Soviet withdrawal has continued to deteriorate over the years, falling into disrepair and disuse. The area that was actually irrigated decreased by almost 50 percent between 1985 and 2006 (FAO, 2009), substantially curbing the viability of higher-value but more irrigation-intensive, crops. Many farming households already cultivate high-value agriculture (HVA) crops such as fruits and vegetables, but in part due to irrigation constraints, they grow them only in small amounts and for household consumption. Grains such as wheat constitute most of the crops produced, but grains have limited commercial viability in Armenia (Fortson, Blair, and Rangarajan, 2010).

The aim of the Millennium Challenge Corporation's Compact with Armenia ("the Compact"), a five-year agreement signed in March 2006, was to increase household income and reduce poverty in rural Armenia through improved performance of the country's agricultural sector. The Compact, managed by the Millennium Challenge Account with Armenia (MCA-Armenia), was originally designed to include two projects: (1) the Rehabilitation of Rural Roads Project and (2) the Irrigated Agriculture Project.[2] The Irrigated Agriculture Project comprised two complementary activities, the rehabilitation of irrigation infrastructure ("the Irrigation Infrastructure Activity", hereafter Infrastructure Activity) and the provision of training, technical assistance, and access to credit for farms and agribusiness ("the Water-to-Market Activity," hereafter WtM Activity). The Infrastructure Activity was intended to provide adequate and timely delivery of water to crop fields, and the WtM Activity was intended to help farmers harness these improvements to introduce new technologies and foster a shift to HVA crop production, both of which would improve household income.[3] The WtM Activity also included technical support to regional water management organizations through the Institutional Strengthening Subactivity (ISSA), with the aim of creating more efficient and consistent

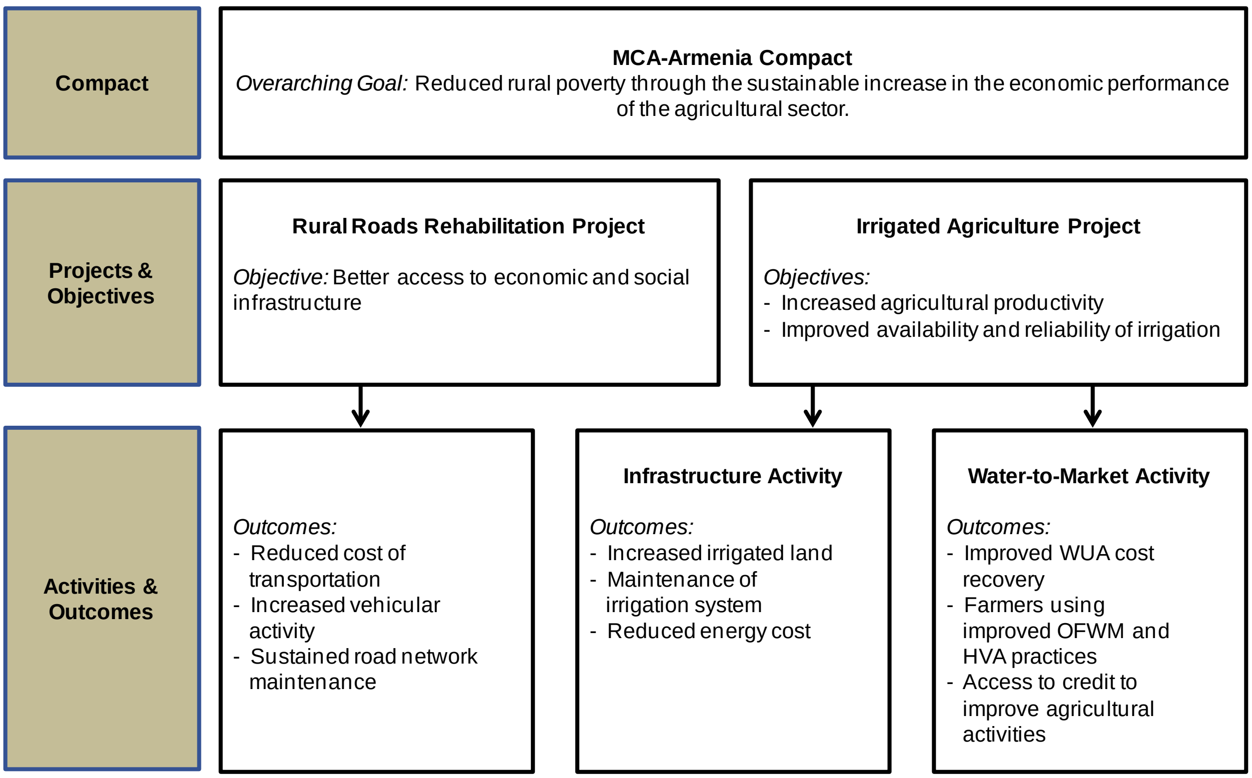
[1] The poverty rate is calculated based on the percentage of households whose consumption per person is less than a threshold that would be required to meet daily caloric requirements.

[2] At the June 2009 MCC Board meeting, the decision was made not to continue funding any further road construction and rehabilitation under the $236 million Compact due to concerns about democratic governance. Approximately 25 km of pilot roads had been completed prior to this decision. Many of the road projects designed with funding from MCC were ultimately financed by the World Bank and the Republic of Armenia.

[3] High-value crops are defined as crops that have relatively high economic value per kilogram, per hectare, or per calorie, such as fruits and vegetables. In Armenia, high-value agriculture consists of all crops that are not grain or grass (Gulati et al., 2005).

irrigation supply and sustaining the investments in irrigation infrastructure. ISSA also included an irrigation policy reform component whereby a reform strategy was developed through a participatory process with stakeholders. By improving living standards among rural residents, these investments were designed to lead to future economic growth in rural areas and throughout the country. Figure I.1 summarizes the overall goal of the Compact and the planned contribution of each activity to the goal.

Figure I.1. Overview of the Compact with Armenia




The Millennium Challenge Corporation (MCC) has commissioned rigorous impact evaluations to examine each of the three main activities of the MCA-Armenia program. This report focuses on the evaluation plans for the Infrastructure Activity. The Infrastructure Activity was completed near the end of the Compact in 2011 and is the final of the three evaluations MCC commissioned. The WtM Activity of the Compact was evaluated previously (Fortson et al. 2013), and the Rehabilitation of Rural Roads Project evaluation report is under review (Fortson et al., forthcoming). The data used in the Infrastructure Activity evaluation cover the 2013 agricultural season, so the evaluation will examine effects approximately two years after completion of the infrastructure work.

In addition to evaluating the Infrastructure Activity itself, the present evaluation will also revisit components of the WtM Activity that were designed to have especially strong complementarities with the investments in irrigation infrastructure. In particular, ISSA was in part intended to enable the water user associations that manage irrigation water to sustain the improved infrastructure with repairs and maintenance. As described previously, WtM also provided farmers with training in on-farm water management and cultivation of HVA, covering agricultural practices thought to be especially valuable to help farmers harness the more reliable water supply from the Infrastructure Activity. ISSA and WtM training were both evaluated as part of the WtM evaluation, but the timeframe preceded completion of the Infrastructure Activity. The present evaluation will provide a longer-term perspective focusing on complementarities that may be realized now that irrigation infrastructure has been rehabilitated.

In the remainder of this chapter we discuss the program logic for the Infrastructure Activity and complementary WtM components, the research questions that stem from the program logic, and how the present evaluation contributes to the existing literature on infrastructure investments in developing countries. In Chapter II we discuss the design for the evaluation of the tertiary canals that were rehabilitated as part of the Infrastructure Activity. In Chapter III we discuss the design for the evaluation of larger infrastructure projects such as main canals and pumping stations. We discuss the follow-up assessment of ISSA in Chapter IV. In Chapter V we discuss the longer-term assessment of adoption of agricultural practices that were introduced as part of WtM farmer training. We cover overarching administrative details of the evaluation in Chapter VI.

## A. Program logic of the Infrastructure Activity and related WtM components

The Infrastructure Activity was implemented by the Irrigation Project Implementation Unit of the World Bank and rehabilitated several different types of irrigation infrastructure, including main canals, the Ararat Valley drainage system, pumping stations, gravity schemes, and tertiary canals. Prior to rehabilitation, water user associations (WUAs), the regional organizations that manage the distribution of and payment for irrigation water in Armenia, estimated that only 25-40 percent of irrigation water actually reached the fields in most of the affected villages. The Compact provided funding of $121 million to rehabilitate irrigation infrastructure schemes across Armenia. The short-term goals of the Infrastructure Activity were to improve the efficiency of irrigation and to increase the area of irrigated land by more than 40 percent (Figure I.2). With access to a more consistent supply of irrigation water, farmers could increase their agricultural production.

In conjunction with the WtM Activity components described below, the Irrigation Project was designed to shift crop production toward higher-value crops, increase sales, and increase agricultural profits for beneficiary farmers. Fortson et al. (2013) describe the WtM Activity more extensively, but here we focus on the specific WtM components that were intended to strongly complement the Infrastructure Activity.

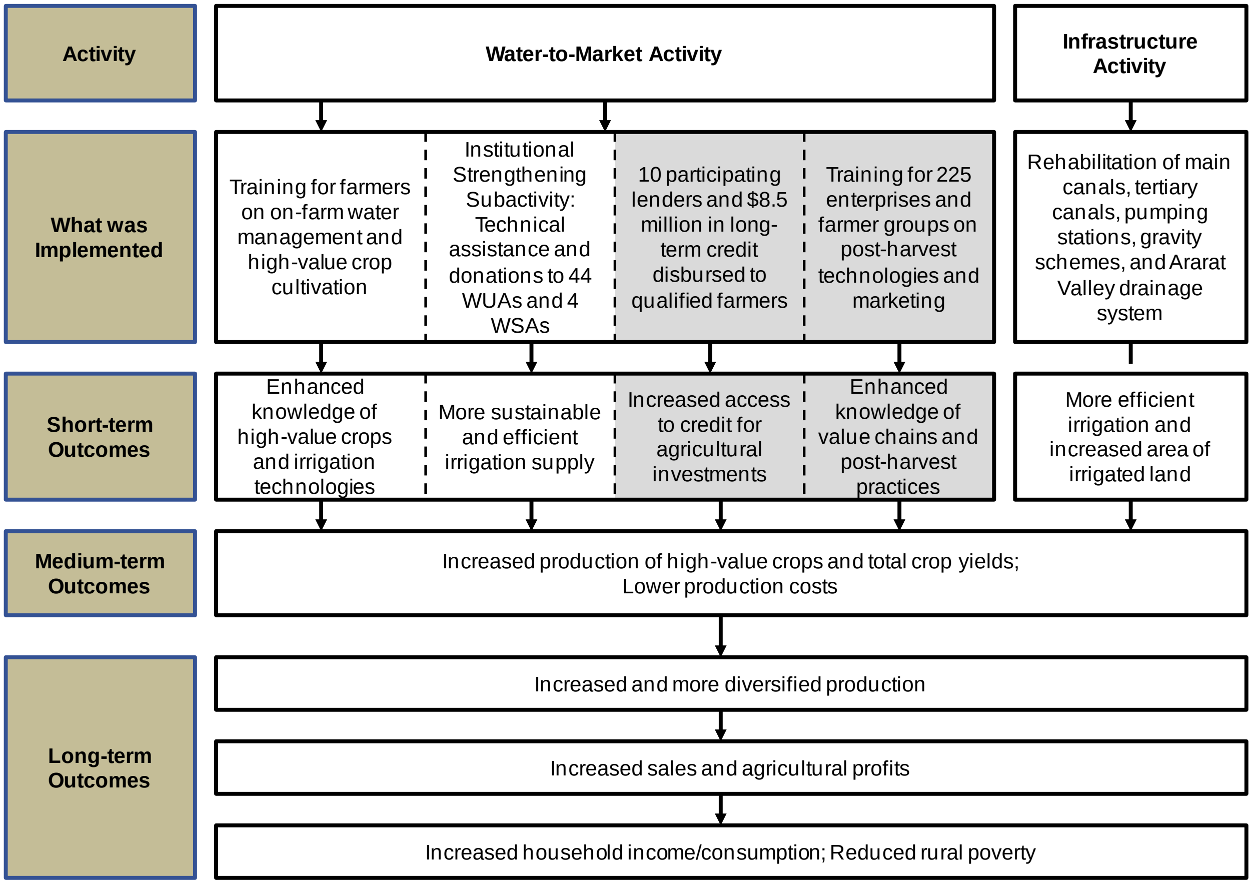
The Institutional Strengthening Subactivity (ISSA), implemented by Mott MacDonald and VISTAA, provided general technical support to WUAs. ISSA also provided assistance to three Water Supply Agencies (WSAs) that operate and maintain irrigation dams and pumping stations. The general aim of ISSA was to strengthen WUAs' and WSAs' managerial, technical, structural, and financial capacity and self-sufficiency. The intent of these improvements was to create more efficient and consistent irrigation supply for WUA members. ISSA also included an irrigation policy reform component, in which a reform strategy was developed through a participatory process with stakeholders.

WtM farmer training included two types of training. On-Farm Water Management (OFWM) training consisted of sessions aimed at helping farmers learn to use new irrigation technologies. As part of this component, demonstration plots were also established to demonstrate the irrigation technologies in practice. A total of 45,000 farmers were scheduled to be trained in

OFWM practices from 2007 to 2010. MCA contracted with ACDI/VOCA and its partners, VISTAA and Euroconsult, (hereafter referred to collectively as ACDI) to implement the training. The goal of this training was for farmers to adopt new and more efficient irrigation techniques, which would lead to increased and more cost-effective agricultural production and higher sales. High-Value Agriculture (HVA) training consisted of establishing demonstration plots and conducting training sessions for farmers on high-value crop substitution and cropping intensity. A total of 36,000 farmers who also received OFWM training were scheduled to be trained by ACDI in HVA from 2007 to 2011. The goal of HVA training was for farmers to adopt new cropping techniques and high-value crops, which would lead to increased and more diverse agricultural production, as well as increased sales.

WtM farmer training in OFWM and HVA was expected to be especially beneficial once farmers had access to improved irrigation. Some farmers who participated in WtM training already had reliable irrigation, but some communities whose irrigation supply was inadequate were provided training with the expectation that the Infrastructure Activity would soon improve their water supply. WtM training began in 2007 and concluded in 2011, but the work was frontloaded, while the Infrastructure Activity began in 2009 and concluded in mid-2011. 2010 was the final agricultural season included in the WtM training impact evaluation, which found that as of 2010, there was no evidence that WtM training had consequential effects on farmers. Agricultural profits and household income were not significantly affected, nor was there evidence of widespread adoption of new agricultural practices that might lead to longer-term impacts in the future. However, it could be that some farmers would invest in new agricultural practices now that irrigation was more readily available in their communities upon completion of the Infrastructure Activity.

Figure I.2. Logic model of the Infrastructure Activity and ISSA




In addition to ISSA and WtM farmer training, the WtM Activity also included a credit component designed to increase access to agricultural loans for farmers, and a component designed to train agribusinesses in post-harvest technologies and marketing. These components are included in Figure I.2 but shaded in gray to indicate that they will not be re-examined in the irrigation evaluation. As indicated in the figure, these components were designed to facilitate the same medium- and longer-term outcomes as the other WtM components and the Infrastructure Activity. The WtM evaluation (Fortson et al., 2013) found suggestive evidence that WtM credit and post-harvest training for agribusinesses may have had effects on the beneficiaries who participated in these components. However, these two WtM components do not have the direct linkages to the Infrastructure Activity that WtM farmer training and ISSA have, nor were there reasons to expect that their effects may have substantially increased over time.

## B. Research questions for the Infrastructure Activity

The overarching research questions guiding the evaluation of the Infrastructure Activity are:

1. Did the program affect the quantity and reliability of irrigation water provided to Armenian farmers?
2. Did farmers adopt new agricultural practices as a result of the program? This question will also assess complementarities between the Infrastructure Activity and the WtM training initiatives.

3. Did the program affect agricultural productivity?
4. Did the program improve household well-being for farmers served by the rehabilitated infrastructure, especially income and poverty?
5. Is there evidence that the infrastructure investments will be sustained after rehabilitation was complete? Are the WUAs themselves financially sustainable?
6. Were the program effects large enough to justify its costs?

Different components of the evaluation focus on different subsets of these questions. The evaluation of the Infrastructure Activity separately examines impacts of tertiary canals and other, larger types of irrigation infrastructure, as discussed in more detail in the respective chapters for each set of irrigation projects. Each of those two (sub-)evaluations focuses on questions 1-4, just for separate types of infrastructure. The reexamination of possible longer-term effects of WtM training provides complementary evidence to help answer question 2 as well. The reexamination of ISSA is designed to help answer question 5, as one of the key purposes of ISSA was to empower WUAs to sustain the irrigation infrastructure investments made under the Infrastructure Activity.

Question 6 will be addressed by reestimating the economic rate of return (ERR) of the activity. In 2008, MCC estimated the ERR for the Infrastructure Activity to be 24.4 percent over a 20-year time horizon (revised from an original estimate of 27.5 percent). This estimate is calculated based on the total costs and benefits associated with the rehabilitation of the canals, gravity schemes, pumping stations, and the Ararat Valley drainage system, as well as the costs of ISSA and MCC's overhead costs, and factors in social benefits and costs as well. The costs for the rehabilitation projects include investment, contingency, maintenance, design, supervision, and other construction costs. The benefits are measured as the estimated incremental agricultural income generated from the improved infrastructure. The calculated ERR is based on a set of assumptions, such as future irrigated land area, cropping patterns, and crop prices, all of which are covered by separate research questions. Thus, in answering the other research questions in the evaluation of the Infrastructure Activity we will assess whether the assumptions in the ERR calculations were accurate, and then we will recalculate the ex post ERR accordingly.

## C. Literature review

Agriculture is a dominant sector in most developing countries, with many rural households relying on agriculture as their primary source of income. Many irrigation systems in the developing world are non-existent or in poor condition, inhibiting farming households from engaging in agricultural production or employment that would improve their well-being. Infrastructure that is either lacking or in poor repair may limit farming households from producing sufficient and high-value crops for their own consumption or for sale in agricultural markets. For these reasons, governments, development banks, and foreign aid agencies have made significant investments to rehabilitate irrigation infrastructure in many developing countries: the Food and Agriculture Organization documented 248 different irrigation infrastructure projects totaling over $8 billion in investment costs from 1980-2000 (FAO, 2000).

The Infrastructure Activity represents a significant investment in infrastructure rehabilitation for Armenia, providing $121 million over the Compact period. In comparison, as of 2005, the operations and management requirements for irrigation water services in Armenia were estimated to be $16 million, half of which was contributed by the government (FAO, 2009). In addition to the physical infrastructure improvements, many countries are also re-considering the way that water resources should be managed to encourage efficiency of water usage, and they are shifting towards a more decentralized system where local water groups assume responsibilities for irrigation operations and maintenance (Hodgson 2007). Our literature review includes studies discussing physical infrastructure improvements relevant to the Infrastructure Activity and irrigation management transfer relevant to ISSA.

Armenia and other post-Soviet republics are distinct from many developing countries in that extensive irrigation systems exist and were once functioning well in the Soviet era, but those systems have largely fallen into disrepair following the collapse of the Soviet Union. Many of the projects included in the Infrastructure Activity were efforts to revive existing infrastructure rather than build new systems. To our knowledge, there have not been rigorous studies of irrigation initiatives in a similar setting as the Infrastructure Activity in Armenia. However, other studies have provided evidence of the effects of irrigation improvements in different settings. These studies generally find that irrigation is associated with a greater level of resources, particularly through higher production and income. One recent empirical review of the literature in Asia shows that across country-specific studies, irrigation is associated with higher cropping intensity, land productivity, and labor employment and wages; irrigated settings also see higher income, lower income inequality, and lower poverty than rain-fed settings (Hussain and Hanjra, 2004). Van Den Berg and Ruben (2006) look at how Ethiopia's national irrigation improvements have affected income inequality by examining ex-post outcomes and find that households with irrigation have higher expenditures and lower dependence on public programs than households without irrigation after accounting for preexisting differences, and with changes in prices, there is an ambiguous effect on inequality. Another study examines how the redistribution of water to canals (through motorized pumps) affects poverty, agricultural production, and nutrition in Northern Mali (Dillon, 2008). Over the eight-year evaluation period, households with this type of irrigation access show higher household consumption, agricultural production, and caloric and protein intake than households without access. They also tend to save more and share more of their resources with fellow village members.

Dams, which are different from the present context in an engineering perspective but nonetheless important in their routing of water to canals (and thus providing access to irrigation), have also featured in the irrigation literature. From the studies reviewed, the welfare impact of dams seems to be ambiguous, yielding economic benefits to some yet inflicting harm on others, including the costs of relocating individuals and environmental costs. In a seminal paper, Duflo and Pande (2007) evaluate the effect of large dams in India and show that agricultural productivity increases but poverty reduction is not universal, and the impacts depend on the proximity to the dam itself. Specifically, agricultural production increases and poverty decreases in downstream districts but there is no significant increase in agricultural production and an increase in poverty in the districts where the dam is built. In fact, on the aggregate level, poverty actually increases. Amacher et al (2004) also find distributional impacts, though for smaller dams in Tigray, Ethiopia. While microdams are found to increase the productivity of fuelwood collection and crop production, villages close to the site of the microdams face a higher rate of waterborne disease prevalence, resulting in more time at home sick or caring for sick family members.

Del Carpio et al. (2011) and IEG (2008) are particularly relevant to the present evaluation because they estimate the impacts of rehabilitating existing irrigation infrastructure, and do so using a comparison group design. Del Carpio et al. (2011) examine the impact of rehabilitating irrigation infrastructure on expenditure, agricultural production, and income ratio measures in coastal Peru. Using national household survey data from 1998-2007, the study identified treatment and comparison groups based on distance to the rehabilitation site. The study shows differential impacts for poor and non-poor farm households. For poor farm households, the infrastructure rehabilitation decreased agricultural production and sales but increased household well-being overall, which they conclude is because poor households substituted work on their own farms for work on others' larger farms. For non-poor farm households, these economic measures increased, but the effects were not statistically significant (or very weakly so).

The IEG (2008) study, commissioned by the World Bank, evaluates command area improvements from both new construction and rehabilitation of existing infrastructure in the Andhra Pradesh region of India. The study finds that there were favorable impacts on yield, cropping intensity, non-farm income, and wage employment over the one-year evaluation period. However, they also show that there was less crop diversification than expected, water wastage in the upper reaches of the canals, and very significant cost overruns and construction delays. Consequently, despite the positive impacts on income, the calculated economic rate of return for the project was just 2 percent, compared to the ex ante estimate of 19 percent. This study illustrates that, considering the cost of most irrigation interventions, examining whether the effects are big enough to justify the costs is crucial, which has received too little attention in the existing literature.

We are unaware of existing literature that attempts to estimate the causal effect of providing technical assistance to WUAs as was done for ISSA. Somewhat greater attention has been given to evaluating the effects of irrigation management transfer, including establishing WUAs, though even this literature is thin (Xie 2007). However, Mukherji et al. (2009) review over 100 case studies of irrigation management transfer in Asia and show that only 43 of 108 were successful according to their definition, with Central Asian countries being particularly weak. To the best of our knowledge, Bandyopadhyay et al. (2007) is the only study to use a comparison group design to measure the impact of irrigation management transfer. They find that irrigation associations with irrigation management transfer show higher frequency of canal maintenance and higher farm productivity, though these findings simply compare post-intervention outcomes for associations with and without management transfer. The IEG (2008) study (discussed above) presented evidence that WUAs in the Andhra Pradesh have limited control over O&M, fee collection, and dispute resolution, and do not greatly empower the poor through participation or leadership. Wang et al. (2006) show that in China, incentives to water managers are more important than farmer participation in water management for water conservation, which in turn is not found to increase poverty or decrease income. A separate study by Wang et al. (2010) shows that WUAs are becoming more common in China and that the World Bank support of WUAs has been successful in terms of participation, reliable water supply, and fee collection but provided no clear benefits in terms of yield, income, and cropping structure. Although these studies of irrigation management transfer do not directly relate to the ISSA intervention, the findings from these studies still highlight the strengths and weaknesses found for WUAs in other settings.

Our study will contribute to the literature by using a rigorous evaluation design to estimate program impacts in a context that has not previously been rigorously examined. We also examine mediating pathways of rehabilitating irrigation infrastructure, which are underexplored in the existing literature, rather than focusing solely on longer-term outcomes such as agricultural production, household income, and poverty. Although the longer-term outcomes are certainly important, examining the mediating pathways helps to understand why any impacts on longer-term outcomes did or did not materialize.

A remaining gap in the literature is a rigorous evaluation of irrigation infrastructure rehabilitation using a randomized-controlled trial. The non-experimental designs used even in the more rigorous studies in the existing literature as well as the present study cannot ensure that the differences in outcomes for the treatment and comparison groups can be attributed to the intervention because not every factor that determines selection of irrigation projects for rehabilitation can be accurately measured and controlled. RCTs have not been used to evaluate irrigation projects mainly because foreign aid agencies understandably choose to fund the projects that are projected to provide the greatest net benefits rather than randomly select projects. Irrigation projects might be selected based on estimated economic rates of return, number of beneficiaries who would be served, or perceived demand for the project. However, considering the frequency and scale of irrigation infrastructure rehabilitation programs, it is our hope that an aid agency will consider using an RCT to rigorously evaluate a future irrigation program, perhaps randomizing among a larger set of irrigation projects that would qualify for funding but cannot all be served within a set budget. Tertiary canal rehabilitation would be an especially suitable choice for evaluation in an RCT because the relatively small size of tertiary canals means many projects are considered under each rehabilitation program, and thus the potential sample size is larger.

# II. EVALUATION DESIGN FOR TERTIARY CANALS

In this chapter we discuss the evaluation design for one subset of irrigation infrastructure projects rehabilitated under the Compact, tertiary canals. Tertiary canals route irrigation water from larger irrigation infrastructure such as main canals or reservoirs to the farmers' fields and are thus vital for farmers. The canals that were considered for rehabilitation deteriorated considerably over the past 20 years, and water losses were substantial. MCA-Armenia estimated that only 25-40 percent of irrigation water actually reached the fields in most of these villages prior to rehabilitation.

Over a hundred communities were initially selected for rehabilitation from amongst those that applied for consideration. Work for most of these began in the first quarter of 2010. MCA-Armenia provided most of the financing for the rehabilitated canals, but villages were responsible for paying a small portion (15 percent) of the construction costs; if they were unable to come up with the co-funding, the canal would not be rehabilitated. This co-funding arrangement was designed in large part so that villages feel ownership over the canals and are more likely to maintain them over the longer term.

Most of the canals selected for rehabilitation are included in the present evaluation. Four were pilot canals that were not considered for the evaluation. Because the pilot canals were rehabilitated well before the other canals and potentially could have been utilized in the previous agricultural season, we could not obtain the informative pre-intervention baseline data that would be necessary for these canals to be included in the evaluation. Others were excluded from the rehabilitation because the community did not contribute the requisite co-funding, so the canal was not rehabilitated.

As detailed in Chapter I, MCA's irrigation rehabilitation efforts covered several different types of irrigation infrastructure, including main canals, the Ararat Valley drainage system, pumping stations, gravity schemes, and tertiary canals. However, for most of these types of infrastructure, only a handful of projects were implemented, too few to support a rigorous evaluation, and there were no other hypothetical projects that could serve as a comparison group. Hence, MCC and Mathematica originally agreed to focus the evaluation effort on the tertiary canal rehabilitation efforts because the prospective research design was rigorous and precise enough to yield informative impact estimates on an important intervention.

Subsequently, MCC asked Mathematica to develop an evaluation design to estimate, as best we can, the impacts of other irrigation infrastructure. The design of the evaluation of other, larger irrigation infrastructure is detailed in Chapter III, as well as the limitations of that evaluation design, one of which is greater risk of biased impact estimates.

Although a random assignment design is considered the most rigorous evaluation approach and may have been feasible in this context, randomly selecting which tertiary canals would be rehabilitated was not done. Communities had to first apply to be considered for inclusion, and then canals were selected based on other factors, particularly engineering considerations and projected economic rates of return. Instead, we have developed a *comparison group* design. Under this approach, tertiary canals for which rehabilitation is planned will be matched to other canals sharing similar geography, pre-rehabilitation conditions, and where similar crops are grown. Examining how outcomes change for farmers in the comparison group, whose canals were not rehabilitated, will inform us about how those outcomes would have changed in the absence of the rehabilitation efforts.

The evaluation of tertiary canal rehabilitation is designed to focus on the following four questions:

1. Did the program affect the quantity and reliability of irrigation water provided to Armenian farmers?
2. Did farmers adopt new agricultural practices as a result of the program?
3. Did the program affect agricultural productivity?
4. Did the program improve household well-being for farmers served by the rehabilitated infrastructure, especially income and poverty?

Additionally, the estimates from this evaluation contribute to answering a fifth research question when examine the benefits and costs of the Infrastructure Activity:

5. Were the program effects large enough to justify its costs?

The remainder of the chapter is structured as follows. Section A describes the comparison group design and how the comparison group was identified. Section B presents the data sources that will be used. Section C outlines the analysis plan. Section D discusses our projections of the statistical precision of the estimated impacts.

## A. Comparison group design

The comparison group design focuses on comparing communities served by rehabilitated tertiary canals (hereafter "tertiary canal communities" or "treatment group") to similar communities whose infrastructure was not rehabilitated (hereafter "tertiary comparison communities").[4] We will estimate the impacts of the program by comparing the post-rehabilitation outcomes for these two sets of communities. Crucially, the analysis will compare how the outcomes have changed relative to the same outcomes measured before the rehabilitation. This approach, which estimates program impacts as the "difference in differences" for the two groups, is stronger than simply comparing post-rehabilitation outcomes because it allows us to adjust for pre-existing differences in the two groups. Still, for this approach to be credible, we must be able to identify communities that are very similar on observable characteristics to serve as the comparison group and properly adjust for remaining differences in observable characteristics.

For a given tertiary canal community, we want to identify a set of communities that, prior to the rehabilitation, is very similar on the characteristics that could be expected to affect the key outcomes: agricultural production and irrigation conditions. Comparison groups are often

[4] Some communities have more than one canal, and the rehabilitated canal serves only a subset of farmers in the village. In these cases, the survey and analysis will focus on farmers served by the rehabilitated canal. In the subsequent discussion, we focus on the illustrative example of a single canal per community for expositional simplicity.

chosen using statistical methods that, for each tertiary canal, would find as close of a match as possible on the many community characteristics that could affect these outcomes. However, relying solely on a statistical matching approach would require a data file containing information such as main crops grown, number of farmers, irrigation sources, etc. for all of the communities in the regions where irrigation projects are planned as well as all communities that could serve as possible comparison communities. Such a data file does not exist and would require considerable effort to create.

Instead, MCA staff who are knowledgeable about the agricultural conditions in these communities worked with Water User Association (WUA) directors to identify an initial set of suitable comparison communities for each tertiary canal. The preliminary set of tertiary comparison communities were selected with a focus on the following three criteria:

1. Be in the same geographic area and served by the same WUA;
2. Have similar pre-rehabilitation irrigation conditions as the communities that will benefit from the rehabilitation project, such as similar water losses and source of irrigation water; and
3. Grow similar crops.

A given tertiary canal community could potentially be matched to multiple tertiary comparison communities if more than one is a good match on the above criteria. We then included all such matches in the survey so as to maximize the sample size and, hence, the statistical precision. In other cases, multiple tertiary canal communities may share a set of tertiary comparison villages if they have similar characteristics. In addition, to get a second assessment of the comparability of these matches, the survey team verified the suitability of each matched comparison community when they went into the field to conduct farmer surveys. They collected independent assessments of the three main criteria listed above from the village mayors as part of a village mayor survey, and they also considered other community characteristics that may indicate that, for a variety of reasons, the planned comparison community did not provide a compelling match. MCA-Armenia also already identified five tertiary canal communities that did not have a suitable comparison community, and those five were excluded from the survey.

The baseline analysis of the TCS (Fortson et al., 2010) found that this process yielded a tertiary comparison group that was observably similar to the tertiary canal communities on most dimensions at baseline. For example, there were no statistically significant treatment-comparison differences in irrigation practices and the cultivation, sales, and value of most types of crops, or average agricultural income and household income. However, there were observable differences on a handful of outcomes at baseline: treatment group farmers were more likely to cultivate grain and their corresponding wheat production was significantly higher; agricultural expenditures were higher among the treatment group than among the tertiary comparison group; and poverty rates were slightly higher for the treatment group than the comparison farmers (although the latter two results were only on the margin of statistical significance). Although these differences were mostly small, they indicate that the tertiary canal communities and tertiary comparison communities were not perfectly matched.

To account for the remaining differences between the tertiary canal communities and the initial set of tertiary comparison communities, we will combine a propensity score weighting approach with regression adjustment. This plan is detailed in Section C of this chapter. We also note that the same methodology is used for the evaluation of large irrigation infrastructure discussed in Chapter III.

## B. Tertiary Canal Survey

The primary data source will be a household survey tailored to this impact evaluation, the Tertiary Canal Survey. The TCS is modeled closely after the survey used for the Water-to-Market impact evaluation, the Farming Practices Survey (FPS), and was fielded by the same survey team led by AREG. As with the FPS, the key outcomes of interest from the TCS include crops cultivated, crop production, agricultural profit, household income, and poverty. The TCS also features questions about reliability and quality of irrigation water. We conducted two rounds of the TCS. The baseline TCS was fielded beginning in December 2009 and finishing in March 2010. The final round was fielded beginning in December 2013 and finishing in March 2014. Thus, the follow-up data were collected four years (and four agricultural seasons) after the baseline and two years after irrigation infrastructure rehabilitation was completed.

Initial plans called for the follow-up data to be conducted one year earlier than was done, but this was moved back by one year to allow for a longer initial follow-up period. Even then, this evaluation may not fully capture all longer-term effects of the tertiary canals. MCC’s ex ante models of the economic rates of return projected that the benefits for tertiary and main canals would increase substantially until about 2018, at which point they plateau. Thus, a limitation of this evaluation is that it might not fully reflect the full long-term benefits of the Irrigation Activity, though it does more accurately measure the intermediate effects that may arise in the medium term, such as transition to higher-value crops.

The sample frame for the TCS comprises the farming households served by the rehabilitated tertiary canals and the matched tertiary comparison group. Prior to the baseline TCS, the survey team worked with village mayors to identify the farmers served by each tertiary canal. Fifteen farmers were selected for interviews in most treatment communities, with rare deviations if a sampled farmer did not show up. Twenty farmers were interviewed in each comparison community. The larger number of respondents in tertiary comparison communities was to allow some cushion in case a few of the comparison group farmers were appreciably dissimilar to the associated treatment group farmers. A total of 2,990 farmers were originally interviewed across 175 communities. For the follow-up survey two communities (one treatment and one comparison) were dropped due to gunfire from across the Armenian border with Azerbaijan. Of the 173 that remain, 89 of these communities are in the tertiary canal treatment group and 77 communities are in the comparison group. The remaining nine communities will be excluded from the tertiary canal evaluation because their canals were ultimately not rehabilitated, as described previously, but we collected follow-up data on them for use in the large infrastructure evaluation. The sample is approximately evenly split between farmers in treatment communities and comparison communities, with about 1,500 in each group. Eighty-three percent of the remaining sample was successfully re-interviewed for the follow-up survey in the 173 communities in which interviews were attempted.

As discussed in Chapter I, the ultimate goal of the MCA-Armenia program is to increase household income in rural Armenia, and hence, these outcomes are an important focus of the TCS instrument. Because a full accounting of all sources of household income would require far longer to administer than the allotted time for each interview, the survey concentrates on sources of income that are most directly affected by irrigation rehabilitation, specifically, income from agricultural production and from employment by the farmer and his or her immediate family. We can also use the average sale price of specific crops for other farmers in the same geographic area to monetize crops that are consumed by the household or bartered. Additionally, the TCS asks for estimates of expenditures on key categories of consumption, and for income from other sources. Table II.1 summarizes the key final outcomes that can be examined using the TCS data. We note that some of these outcomes, such as employment income or income from pensions and other sources, are not outcomes the program is intended to directly affect. However, we include them because they are important components of the household's total income and, hence, it is necessary to have estimates of them. Rehabilitating infrastructure may change the economic returns for different activities, inducing farming households to substitute away from outside employment in favor of more time for agriculture. Conversely, as was found by Del Carpio et al. (2011), households with smaller farms may find that employment opportunities are better on large farms following rehabilitation, in which case they might substitute away from work on their own farms in favor of working on others' larger farms.

Table II.1. Final outcome measures: survey data

| Final outcome measures | Time frame |
|---|---|
| ***Agricultural Productivity.*** Total amount of specific crops grown per household; amount of crops grown per square meter; total value of all crops cultivated per household. | Last Agricultural Season |
| ***Livestock.*** Number of cows, pigs, and sheep owned. | As of Survey Date |
| ***Revenue from Agricultural Production.*** Value of crops sold; total value of all crops (including those sold, bartered, or consumed). | Last Agricultural Season |
| ***Agricultural Costs.*** Expenditures on fertilizers, pesticides, irrigation water, hired labor, rented equipment, and taxes. | Last Agricultural Season |
| ***Profit from Agricultural Production.*** Revenues minus costs—the income from agricultural activities. | Last Agricultural Season |
| ***Income from Employment.*** Whether household head, spouse, and any grown children were employed (besides work on the family farm); total earnings from employment. | Last Year |
| ***Income from Pensions, Remittances, or Social Programs.*** | Last Year |
| ***Total Household Income.*** Agricultural profits plus income from employment or other sources. | Last Year |
| ***Household Consumption.*** Expenditure on purchased food, health care, housing products, utilities, and transportation; cost of purchased goods plus value of crops consumed by the household. | Typical Month/Last Year |

Although examining impacts on the key outcomes shown in TableII.1 is valuable, the impact estimates may not capture the full effect of the irrigation rehabilitation on household well-being in this timeframe for two reasons. First, some farmers may be slow to adapt to the improved irrigation conditions, and those changes would not be observed in the two-year window between when infrastructure rehabilitation was completed and follow-up data were collected. Second, changes may lead to improvements only after a longer time horizon. For example, if improved irrigation allows farmers to plant orchards in fields that are currently fallow, those orchards might not be mature enough to bear fruit in the first two agricultural season. Thus, examining intermediate outcomes will be especially crucial for this impact evaluation so that we can gauge whether future improvements in household well-being are possible. We would expect an impact on households' income only if we observe that a substantial proportion of the targeted farmers actually had improved irrigation, and perhaps most importantly, are then utilizing the improved irrigation to improve their agricultural productivity. Indeed, MCC modeled the primary paths for increasing farm income through an economic rate of return model (ERR) that justified the project for funding. The ERR included two agricultural benefits as a result of improved irrigation infrastructure. Those were (1) increased land under high-value agriculture (for example, switching crops from wheat to cucumbers) and (2) higher yields. Examining the intermediate outcomes also establishes the counterfactual—what the irrigation conditions would have been even in the absence of irrigation rehabilitation. Table II.2 summarizes the key intermediate outcomes that can be examined using the TCS data.

We note that the question about perceptions of other households' maintenance of irrigation infrastructure is an outcome that has not been previously measured in the TCS. It was included this year to help assess sustainability of the infrastructure investments and to assess a particular concern that some farmers counteract the rehabilitation efforts by, for example, putting old cars in canals to raise water levels or disposing of trash in canals. This behavior has been observed for the existing infrastructure.

Table II.2. Key outcome measures: survey data

| Intermediate outcome measures | Time frame |
|---|---|
| ***Water Usage.*** Amount of land that could be irrigated; amount of land that actually was irrigated; amount of land watered using other sources (such as well or drinking water); frequency of irrigation; estimated amount of irrigation water used. | Last Agricultural Season |
| ***Quality of Irrigation System.*** Perceived overall quality of irrigation in the village; perceived changes in quality from previous year; main irrigation problems in the village; timeliness and sufficiency of irrigation water. | Last Agricultural Season |
| ***Perceptions of Community Farmers' Contributions to Maintenance of Irrigation Infrastructure.*** Whether farmers actively help maintain infrastructure or take actions that are detrimental to long-term functionality, such as disposing of trash in canals. | Last Agricultural Season |
| ***Investment in Agricultural Technology or Equipment.*** Ownership of personal reservoir or water pump; irrigation technologies used; WUA membership; use of other efficiency-enhancing or environmental practices. | Last Agricultural Season |
| ***Cropping Patterns.*** Specific crops grown, especially high-value crops; average amount of land devoted to cultivation of each crop; average hectares of land devoted to crops; average amount of land irrigated for each crop. | Last Agricultural Season |

A short survey of village mayors was fielded in tandem with the TCS. The survey asked the mayors in these communities simple questions about what irrigation infrastructure was rehabilitated since 2009 and by whom. The primary purpose of these questions is to provide context to the quantitative findings. In particular, we want to understand what other rehabilitation projects were implemented in the tertiary canal and tertiary comparison communities that could influence any observed impacts. The survey of village mayors also asks mayors whether there were any major events that affected agricultural production, such as droughts or heavy rains.

## C. Analysis plan

Because the communities where infrastructure was rehabilitated were not randomly selected, the tertiary canal and tertiary comparison communities differed in measurable ways before any infrastructure was rehabilitated (Fortson et al., 2010). The baseline TCS, along with community-specific information from the accompanying survey of village mayors, provides crucial data on the key outcome measures prior to the intervention. In this section we discuss the statistical adjustments we will use to account for differences in baseline measures for the treatment and comparison groups.

We use a regression framework combined with weighting communities based on their estimated propensity scores. The key concept underlying propensity score methods is to reconstruct the tertiary comparison group to look similar on observable characteristics to the treatment group. A propensity score is the probability that a community received an irrigation infrastructure improvement. These probabilities can be estimated using household and community-level variables, including variables where the treatment and comparison groups are not similar at baseline.

Conceptually, regression and propensity score methods attempt to do the same thing—account for observable baseline differences—and the key underlying assumption for both methods is that all crucial baseline differences between the treatment and comparison groups are measured in the data. However, each method also depends on correctly specifying the econometric model to account for those baseline differences. Combining inverse probability weighting (IPW) on the propensity score with regression adjustment means the estimates are robust to either misspecification in the propensity score or misspecification in the regression model. Consequently, this method is sometimes referred to as being “doubly robust” (Schafer and Kang, 2008).

We will use this method by applying the following process:

1. **Identify characteristics that will be used to calculate propensity scores.** We will use a saturated model that considers all baseline variables that are possibly correlated with receiving the intervention. Because whole communities are selected for the intervention, the propensity score model will use community-level data. As described above, some variables are already measured at the community level, such as region indicators (Armenian “marzes”), number of households in the village, number of available livestock, percentage of households that farm as their main occupation, total land cultivated, number of hectares cultivated for three most common crops, village is a WUA member, village has an irrigation network, and indicators for other infrastructure that is rehabilitated besides tertiary canals. However, most of the baseline data come from the baseline TCS, and for these we will construct community-level averages. Household-level measures will include characteristics such as: head of household’s age, sex, and education; number of adults and number of children; hectares of arable land, orchards, and vineyards for the household; agricultural

production, revenues, and profit; WUA membership; use of irrigation water; and perception of the condition of the irrigation system.

2. **Calculate propensity scores for the analysis sample.** We will use a logistic regression to estimate a model with the variables from step 1 as the independent variables and a binary indicator variable for treatment status as the dependent variable. Using this estimated model, we will then estimate the propensity score $p_c$ for each community as the predicted value based on the estimated model parameters and the community's baseline variables. We will also assess whether there are any communities with particularly low or high estimated propensity scores, and whether these communities should remain in the analysis sample. Communities with estimated propensity scores that are off the shared "common support" of the treatment and comparison groups—that is, treatment communities that are unlike any comparison group community and comparison communities that are unlike any treatment community—would be excluded.

3. **Construct inverse probability weights (IPW) using the propensity score.** We will next use the estimated propensity scores to reweight the comparison group communities to be observably similar to the treatment communities (apart from any communities excluded in step 2). Each tertiary comparison community will be assigned a weight of $p_c/(1 - p_c)$. Intuitively, tertiary comparison communities that look most similar to treatment communities ($p_c$ closer to 1) receive more weight than communities that look less similar ($p_c$ closer to 0). Each treatment community will be assigned a weight of 1. These weights are constructed so that we are estimating the average impact of treatment for the communities in the treatment group.

4. **Assess baseline differences using IPW.** We will then use the weights to calculate weighted differences in all key baseline characteristics. If the differences are sizeable, then we will repeat steps 1 and 2 with expanded sets of covariates, including interactions of variables and the inclusion of higher-order terms for variables most predictive of treatment status.

The impacts of the tertiary canals will be estimated based on outcomes measured in the follow-up TCS, two years after the irrigation projects are complete. The impact of tertiary canals on a given outcome measure will essentially be calculated by subtracting the average value of that outcome measure for the comparison group from the average value for the treatment group. As noted above, it is crucial to adjust for pre-existing differences between the treatment and comparison groups to ensure any observed differences at the time of the follow-up survey can be credibly attributed to the program. We will use a regression framework with the propensity score weights described above. Using regression models to control for these baseline characteristics also improves the statistical precision of the impact estimates. The basic regression model can be expressed as follows:

$$(1) \quad y_{iv,F} = \beta' x_{iv,B} + \gamma T_v + \eta_v + \varepsilon_{iv},$$

where $y_{iv,F}$ is the outcome of interest for household $i$ in village $v$ at the follow-up survey; $x_{iv,B}$ is a vector of baseline characteristics; $T_v$ is an indicator equal to one if village v is in the treatment group and zero if it is in the comparison group; $\eta_v$ is a village-specific error term; and $\varepsilon_{iv}$ is a random error term for the household. The parameter estimate for $\gamma$ is the estimated impact of the program.

The vector of baseline characteristics $x_{iv}$ will include both household and village characteristics. We will control for village characteristics such as the geographic region, size of the community, and number of farmers as well as whether the village was a beneficiary of WtM training or rehabilitation of other irrigation infrastructure. We will also control for household size and composition, and characteristics of the household head, namely, education level, gender, age, and number of years farming. Baseline measures of the outcome measures of interest will also be included in the regression. The regression models must also account for the fact that, because farmers served by the same canal are exposed to the same effects of weather and other idiosyncratic shocks, their outcomes will be correlated and cannot be considered statistically independent. This "clustering" of farmers is reflected in the village-specific error term $\eta_v$. Lastly, we will use bootstrapped standard errors to account for estimation error in the propensity score.

An alternative specification of (1) would define $y_{iv,F}$ as the growth in outcomes, rather than the post-rehabilitation measure of the outcome. This formulation is sometimes preferred in situations where the outcome is measured with error due to recall error, which is usually the case for complicated outcome measures such as household income. Hence, we plan to conduct sensitivity analyses using this alternative specification.

In addition to the quantitative impact evaluation described in this chapter, MCA-Armenia also funded a qualitative process analysis that investigated issues such as how the irrigation rehabilitation project was designed, the fidelity of program implementation, and stakeholders' perceptions of program implementation and benefits. This qualitative information will provide valuable insights that complement the quantitative findings by helping us determine why the expected program impacts did or did not occur.

We will explore whether the impacts of the tertiary canal rehabilitation differ for two subgroups of beneficiaries: female-headed households and households that were living in poverty at baseline. The analysis would essentially estimate separate impacts of irrigation for the subgroup. The impacts would be estimated by modifying equation (1) as follows:

$$(2) \quad y_{iv,F} = \beta' x_{iv,B} + \gamma_{S=1} T_v \times S_{iv} + \gamma_{S=0} T_v \times \left(1 - S_{iv}\right) + \eta_v + \varepsilon_{iv},$$

where $S_{iv}$ equals 1 if the household belongs to the specified subgroup (for example, female-headed households) and 0 if it is not (for example, a male-headed household rather than female). Testing whether $\gamma_{S=1}$ differs from $\gamma_{S=0}$ would tell us whether the impacts differ for the two groups. Analogously, we could estimate whether there irrigation impact differs for communities that did or did not also have rehabilitated large infrastructure.

## D. Minimum detectable impacts on key outcomes

In this section we present estimated minimum detectable impacts for the evaluation of tertiary canals. Minimum detectable impacts are the smallest true program effects that can be reliably detected given the sample sizes and methodology. If the true program effect is smaller than the estimated minimum detectable impact, then it is unlikely that the evaluation will find a statistically significant effect.

Minimum detectable impacts depend on many parameters. Some parameters are known or can be controlled, such as the sample size and whether households in the treatment group are assigned individually or by community, as they are in the present evaluation. Many parameters must instead be projected, drawing on prior research whenever possible, such as the variance of the key outcome measures, the extent to which controlling for key baseline covariates reduces variability in those key outcomes, and the degree of within-group clustering in the case of community-level assignment. Because the present study is structured similarly to the impact evaluation of WtM training and uses similar data (Fortson et al., 2013), we assume that all parameters will be similar except for the sample sizes and any design effects introduced by using propensity score weighting.

Based on the similarity of the tertiary canal treatment and tertiary comparison groups at baseline, we expect that the propensity score weights will be mild, which means weights are not likely to affect the standard errors for the tertiary canal evaluation. However, given the smaller sample size for the tertiary canal evaluation compared to the WtM training evaluation, particularly the number of communities included, we project that the standard errors for the impact estimates will be approximately 15 percent greater. The minimum detectable impacts will correspondingly be 15 percent greater as well. Table II.3 reports minimum detectable impacts for four illustrative outcomes.

Table II.3. Minimum detectable impacts of tertiary canals on key outcomes

| | Estimated minimum detectable impact | Baseline average |
|---|---|---|
| Irrigated Land (hectares) | 0.14 | 0.77 |
| Land Under Cultivation for HVA Crops (hectares) | 0.09 | 0.61 |
| Agricultural Profits (USD) | 314 | 1,286 |
| Economic Income (USD) | 482 | 4,684 |
| Households Below the Lower Poverty Line (%) | 6.0 | 16.0 |

Source: Authors' calculations drawing on Fortson et al. (2013), which used the Farming Practices Survey, a survey with outcomes and sample that are closely related to the TCS used for the present analysis.

Note: The minimum detectable impacts assume a confidence level of 95 percent, two-tailed tests, and 80 percent power, resulting in a factor of 2.8. The minimum detectable impact uses the projected standard error multiplied by this factor. The projected standard error is calculated using the estimated standard errors in Fortson et al. (2013) multiplied by 1.15 to account for the smaller sample size in the present analysis.

# III. EVALUATION DESIGN FOR LARGE INFRASTRUCTURE PROJECTS

Most of the irrigation infrastructure projects that were rehabilitated under the Infrastructure Activity are large projects that affect several communities. For example, each main canal serves at least 15 communities and often much more. Most pumping stations serve one to five communities, but some serve a dozen or more. As described in Chapter II, the evaluation of the Infrastructure Activity originally focused on the rehabilitation of tertiary canals because the evaluation would answer questions relevant for all irrigation projects and a rigorous evaluation design. However, considering the amount of funding dedicated to the larger infrastructure and the number of beneficiaries, MCC subsequently decided to conduct analyses of the effects of rehabilitating the other irrigation structures to the extent possible using existing data sources.[5]

Apart from tertiary canals, each type of irrigation project (main canals, pumping stations, gravity schemes, and drainage system) covers too few communities for us to estimate separate effects for these projects. Instead, we pool them together to estimate impacts of having larger infrastructure rehabilitated.

The evaluation design for estimating impacts of larger infrastructure shares many features with the tertiary canal evaluation described in Chapter II. Most prominently, we determined that the TCS, which was originally designed exclusively for the tertiary canal evaluation, was the data set that was most appropriate for the evaluation of larger infrastructure as well.[6] We also use a comparison group evaluation design, comparing communities served by any large project (hereafter "LI treatment communities," where LI designates it is the treatment group for the large infrastructure evaluation) to other communities whose infrastructure was not rehabilitated (hereafter "LI comparison communities"). We will estimate the impacts of the program by comparing the post-rehabilitation outcomes for these two sets of communities, using propensity score weighting and regression adjustment to account for pre-existing differences. We note that the LI treatment communities are not the same set of communities as the tertiary canal communities described in Chapter II, nor are the LI comparison communities the same as the tertiary comparison communities.

The evaluation of large infrastructure rehabilitation is designed to focus on the same set of research questions as the tertiary canal evaluation, just for a different set of infrastructure:

1. Did the program affect the quantity and reliability of irrigation water provided to Armenian farmers?
2. Did farmers adopt new agricultural practices as a result of the program?

---

[5] Because baseline data would be crucial for any nonexperimental design, and because it was too late to conduct a pre-intervention baseline survey, it was not possible to tailor a survey to the evaluation of larger infrastructure.

[6] We considered data sources that are being used for evaluations of other Compact activities, either as alternatives or supplements to the TCS, but each was found lacking for one or more reasons. These other data sources that were considered include the Farming Practices Survey (final round completed before large irrigation projects are complete), the Water Users Survey (cross-sectional, and thus less useful econometrically), and the Integrated Living Conditions Survey (cross-sectional, and final round will be completed concurrently with the completion of large irrigation projects).

3. Did the program affect agricultural productivity?
4. Did the program improve household well-being for farmers served by the rehabilitated infrastructure, especially income and poverty?

Additionally, the estimates from this evaluation contribute to answering a fifth research question when examine the benefits and costs of the Infrastructure Activity:

5. Were the program effects large enough to justify its costs?

In the remainder of Chapter III we discuss our evaluation plan for large irrigation infrastructure. Because so many features of the large irrigation infrastructure mirror our methodology for the tertiary canal evaluation, we focus on the differences between the two, and refer the reader back to Chapter II for details that are the same in both evaluations.

## A. Tertiary Canal Survey

In the evaluation of large infrastructure, we will estimate impacts on the same set of outcomes that are measured in the TCS for the tertiary canal evaluation as described in Chapter II. Likewise, we include all 173 communities that were interviewed in the follow-up round of the TCS, though the distribution of treatment and comparison communities is different from the tertiary canal evaluation. For the large infrastructure evaluation, 107 communities are in the LI treatment group and 66 are LI comparison communities (Table III.1). Ideally, we would like to estimate separate impacts for each type of irrigation project, for example, separate impacts for pumping stations, main canals, etc. However, the number of communities benefitting from each type of large irrigation project is generally small, and the potential comparison group for each separate type of project is even smaller. Hence, our analysis pools all of these communities and is designed to estimate the impact of a large irrigation project rehabilitated, regardless of which type.

Table III.1. Number of treatment and comparison communities included in TCS sample

| Project type | TCS communities served |
|---|---|
| Any Large Irrigation Project [a] | 107 |
| Pumping Stations | 39 |
| Drainage | 25 |
| Gravity Scheme | 13 |
| Main Canals | 76 |
| LI Comparison Group | 66 |

[a] Many communities are served by multiple large irrigation projects, so total communities served by any project does not equal the sum of communities served by each type of project.

## B. Analysis plan

The methodology for the large infrastructure evaluation will be analogous to the methodology described in Chapter II for the tertiary canal evaluation. As with the tertiary canal evaluation, we will construct weights based on estimated propensity scores to construct a comparison group that is observably similar to the LI treatment group, and we will use regression adjustment to further account for observable differences. However, there are three key limitations that are unique to the large infrastructure evaluation and will affect how we interpret the impact estimates.

The first issue is that, for the large infrastructure evaluation, the TCS should be considered a sample of convenience. As described in Chapter II, the TCS sample was constructed to estimate average treatment effects for the communities whose tertiary canals were rehabilitated (with the exception of four pilot canals). In contrast, the TCS sample is not a representative random sample of all communities who stood to benefit from large infrastructure projects. That said, the sample includes communities in 9 of the 10 regions of Armenia (excluding the capital city, Yerevan) and all types of infrastructure that were rehabilitated, so we still believe the findings will be informative even if they do not generalize to all beneficiary communities in a purely statistical sense.

The second issue is that the LI treatment and LI comparison communities are more dissimilar at baseline on observable dimensions than are the tertiary canal treatment and comparison communities. Consequently, the estimated propensity scores will be more dissimilar. This means that we are likely to find that more LI treatment communities do not look like any LI comparison communities (and vice versa)—in statistical terms, they are not in the common support of the estimated propensity score distribution—and should be dropped from the analysis. It also means that the remaining weights will be more variable. Both of these issues mean the impact estimates will have less precision, which we attempt to account for when we discuss the minimum detectable impacts in Section C.

Finally, the way we think about the clustering of households for the regression models is somewhat different. Ideally, we would account for clustering at the project level rather than the community level, but several factors render accurate higher-level clustering impossible. First, we do not know which comparison communities would have been grouped together if the comparison communities were to also benefit from large irrigation projects, so we cannot cluster LI comparison communities in the same way as LI treatment communities. Second, several of the large irrigation projects overlap, so the true clustering structure cannot be readily modeled even for the LI treatment group. Third, for some projects individual communities could be selected for the project based on the specific scope of rehabilitation. For example, gravity schemes could be constructed to affect more or fewer communities, or different parts of main canals could be rehabilitated that would benefit different subsets of communities. That we cannot account for higher-level clustering means the standard errors may be understated, and statistical significance may be overstated.

This concern is mitigated somewhat by controlling for geography in the form of marz (Armenian region) fixed effects, which should account for some of the cross-community correlation. Additionally, we can explore the extent to which the standard errors are understated by examining the degree of clustering in a subsample of communities, specifically, the treatment communities without project overlap. With this subsample, we can calculate a design effect, that is, how large the standard errors would be if we were to (correctly) account for higher-level clustering in our model rather than community-level clustering.

## C. Minimum detectable impacts on key outcomes

In this section we present estimated minimum detectable impacts for the evaluation of large infrastructure, analogous to those presented for the tertiary canal evaluation. As described in the previous section, we expect that the treatment and comparison communities for the large infrastructure evaluation will be more dissimilar than was the case for the tertiary canal communities. This means we will need to exclude those communities that are most dissimilar, and the propensity score weights will vary more. We expect this to result in larger standard errors for the large infrastructure impact estimates than the tertiary canal impact estimates' standard errors. Although we cannot confidently say how much larger, we assume that the minimum detectable impacts will be 10 percent greater than for tertiary canals. Specific estimates are reported in Table III.2.

Table III.2. Minimum detectable impacts of large infrastructure on key outcomes

| | Estimated minimum detectable impact | Baseline average |
|---|---|---|
| Irrigated Land (hectares) | 0.15 | 0.77 |
| Land Under Cultivation for HVA Crops (hectares) | 0.10 | 0.61 |
| Agricultural Profits (USD) | 345 | 1,286 |
| Economic Income (USD) | 530 | 4,684 |
| Households Below the Lower Poverty Line (%) | 6.6 | 16.0 |

Source: Authors' calculations drawing on Fortson et al. (2013), which used the Farming Practices Survey, a survey with outcomes and sample that are closely related to the TCS used for the present analysis.

Note: The minimum detectable impacts assume a confidence level of 95 percent, two-tailed tests, and 80 percent power, resulting in a factor of 2.8. The minimum detectable impact uses the projected standard error multiplied by this factor. The projected standard error is calculated using the estimated standard errors in Fortson et al. (2013) multiplied by 1.265—or equivalently, the standard errors used in Table II.3 multiplied by 1.10—to account for the smaller sample size in the present analysis and greater design effects to weighting on the propensity score.

# IV. FOLLOW-ON ANALYSIS OF THE INSTITUTIONAL STRENGTHENING SUBACTIVITY

Created by the Armenian government in 2002, Water User Associations (WUAs) are organizations established by water users to carry out the operation and maintenance of the country's rural irrigation systems. WUAs are nonprofit legal entities that operate in the public interest, often with large government subsidies to cover operational costs. Water Supply Agencies (WSAs) handle the operation and maintenance of irrigation dams and pumping stations and supply water to WUAs. WUAs pay WSAs based on their projected water usage.

The primary objective of the Institutional Strengthening of Irrigation Management Entities Subactivity (ISSA) was to improve the managerial, technical, structural, and financial capacity of WUAs (and WSAs) operating in rural Armenia. According to the ISSA design, WUAs' enhanced capacity would allow them to manage irrigation systems more efficiently and autonomously, and eventually reach financial sustainability. In addition, strengthened WUAs could more effectively operate and maintain Armenia's rural irrigation infrastructure, thus ensuring reliable water supply and supporting long-term rural agricultural development. To meet these multiple objectives, the component's implementing organizations, Mott MacDonald, Euroconsult, and VISTAA, provided technical assistance to staff from all 44 WUAs in Armenia (as well as 3 WSAs) on irrigation water delivery services, water service fee collection practices, budgeting and accounting processes, irrigation infrastructure maintenance, and participatory management principles. ISSA's implementing organizations also provided material assistance to WUAs and WSAs in the form of office equipment, computer software, and heavy machinery. With a budget of approximately $4.9 million, this component was launched in September 2008 and completed in October 2011.

As part of the evaluation of WtM, Fortson et al. (2013) used qualitative data and a short time series (three years, 2008-2010) of WUA administrative data to assess the effects of ISSA on the aforementioned irrigation-related outcomes. As part of the present evaluation we will revisit the WUA administrative data for three additional years (2011-2013) to examine whether there is evidence of longer-term sustainability for WUAs financially and for the irrigation investments that were funded by MCC. We plan to supplement the administrative data with a short set of questions that will ask WUA directors for summaries of their WUAs' repairs to, upgrades of, and maintenance of the rehabilitated infrastructure. We also plan to ask about the extent to which farmers contribute to maintenance (or hinder it), a question that was asked of farmers in the TCS as well. The research questions we will answer for this component are thus:

1. Are the WUAs financially sustainable, that is, how do their revenues compare to their costs? Has this changed over time?
2. Is there evidence that the infrastructure investments will be sustained after rehabilitation was complete? In particular, are WUAs maintaining the infrastructure appropriately, and is the infrastructure being mistreated by farmers?

ISSA included some other components that were covered in the evaluation of WtM that will not be revisited as part of the present report. In particular, Mott MacDonald developed a national policy paper for the Armenian irrigation sector. This paper became the basis for the irrigation reform strategy developed by AVAG Solutions, modified through a participatory process with stakeholders and approved by MCA-Armenia's governing council. WUA and WSA staff also had the opportunity to participate in study tours of irrigation systems in the United States and Europe. Conducted in 2010 and 2011, these tours provided WUA staff and government officials with an opportunity to observe effective and entrepreneurial WUAs, as well as highly functional rural irrigation systems.

## A. Reassessment of WUAs' financial sustainability

As was done in the WtM evaluation, we will rely mainly on WUA administrative data in reassessing ISSA. AVAG Solutions administered this survey to administrative staff in all 44 WUAs served by ISSA. The survey covers the following domains: WUA characteristics, infrastructure and technical capacity, human resources, office space and equipment, water intake and delivery, WUA finances, and institutional arrangements. The survey was already administered in person to WUA staff in each year from 2008-2012 and will be administered again in 2014 to cover the 2013 agricultural season. The years 2008-2010 were covered in the WtM evaluation; this reassessment will add three new years of data but otherwise employ the same simple methodology and set of outcomes.

We do not have a credible way to measure the counterfactual—what would have happened to WUAs in the absence of ISSA—so the evaluation should be considered a performance evaluation rather than a rigorous impact evaluation. To explore possible effects of ISSA on cost recovery, we will use data from the administrative data to measure how the 44 beneficiary WUAs' reported expenditures and revenues changed from 2008 through 2013. A positive change in WUAs' average net revenues from 2008 to 2013 cannot be interpreted as a direct result of ISSA as other changes, such as government regulations or changes in WUA leadership, might have influenced observed outcomes. Nonetheless, measuring the average change across WUAs will offer insight into their potential to achieve long-term financial sustainability, a key objective of ISSA.

## B. Assessing sustainability of irrigation infrastructure investments

In addition to collecting administrative data, we will also collect information on a new and more subjective set of questions asking WUA directors about the functionality of the rehabilitated systems, WUAs' self-initiated rehabilitation efforts, farmers' contributions to maintenance, and WUAs' management practices. These questions focus on the long-term sustainability of the irrigation investments to help in projecting whether any positive impacts that we observe seem likely to persist into the future, and also the financial sustainability of the WUA itself. Tentatively, we plan to ask these questions of all WUA directors in a separate module administered by AVAG Solutions in concert with the administrative data collection. However, if this proves infeasible, we may instead administer it separately for a subsample of the WUAs. The specific topics covered in this module are summarized in Table IV.1.

Table IV.1. Measures of irrigation infrastructure maintenance and WUA financial sustainability

| Outcome measures | Time frame |
|---|---|
| ***Perceptions of Functionality of Systems.*** Whether the WUA director thinks the rehabilitated infrastructure has worked well overall. | As of Survey Date |
| ***Changes in Repair Costs.*** Whether MCA-funded rehabilitation has contributed to reduced repair costs (because infrastructure is in better condition) or increased repair costs (because WUA has more responsibility to maintain it). | As of Survey Date |
| ***WUA-funded Upgrades to Infrastructure and Equipment.*** Whether the WUA has invested in upgrades to irrigation infrastructure, equipment, or computer systems after the Compact ended. | Since 2011 |
| ***Assessment of Farmers' Contributions to Infrastructure Functionality and Maintenance.*** WUA director's assessment of the extent to which farmers take actions that harm the system's functionality, such as disposing of trash or rigging systems to extract water in ways that harm the system's functionality. | As of Survey Date and Also Since 2011 |
| ***Financial and Maintenance Documentation.*** Whether the WUA has developed, uses, and updates a business plan; whether the WUA adheres to the operations and maintenance manual for rehabilitated infrastructure; whether the WUA adheres to the management improvement plan developed with MCA funding. | |

Because these data will only be collected at endline, the analysis of irrigation infrastructure will only be descriptive, and it will be subject to measurement error due to recall issues. However, we believe it will provide valuable insights about the long-term prospects for the Infrastructure Activity.

# V. WTM TRAINING FOLLOW-UP

An important but disappointing finding from the WtM evaluation was that there was little evidence that WtM training had led to widespread adoption of new agricultural practices (Fortson et al., 2013). Two possible explanations for this disappointing result are (a) the follow-up period of two to three years after farmers were first trained was too short, or (b) many farmers did not have access to reliable irrigation water at the time of training, which preceded completion of the Infrastructure Activity, and they could not effectively adopt new practices until the irrigation systems were rehabilitated.

We will conduct analyses to explore these two hypotheses using an opportunistic subsample from the TCS. The questions we seek to explore in these analyses are:

1. Is there an increase in the use of improved practices five to six years after training, especially in communities with rehabilitated infrastructure?
2. If we observe any increases in adoption, are these attributable only to the Infrastructure Activity or is it the interaction between the infrastructure rehabilitation and the WtM training provided earlier? That is, was training effective but only after irrigation was newly available, or was the increase in adoption unrelated to training?

## A. Analytical approach

We will use simple descriptive analyses to answer the first question above. We will tabulate agricultural practice adoption rates for households that are in a WtM treatment village that also received an irrigation infrastructure improvement (tertiary canal or large infrastructure). Many of the practices reported in the TCS were measured at follow-up but not at baseline for this survey, so we cannot look at changes over time for these. Instead, we will report adoption rates at follow-up for these outcomes and refer to the WtM evaluation report to put into context whether these adoption rates are substantively different. We note that this subset of communities is not a representative sample of communities that were selected for WtM training. Rather, it is an opportunistic sample of those communities. Still, it will provide a general sense of whether adoption has greatly increased for WtM communities that were offered training and whose irrigation infrastructure was later rehabilitated. Adoption rates were low enough in 2010 for most of the key practices that it should be evident if there has been widespread adoption in the years since the WtM evaluation was conducted. The final round of the TCS also added a question about farmers' experience with, or fear of, personal irrigation-related equipment being stolen from fields to help further understand why adoption rates are not higher.

The second question listed above can be answered more rigorously, though it will only be meaningful if there is evidence of consequential adoption rates since 2010. As described in Chapter II for evaluating tertiary canals, and analogously for larger infrastructure in Chapter III, we will first estimate separate impacts of tertiary canals and larger infrastructure on adoption of new agricultural practices. This first step tells us whether there was increased adoption in the communities with newly reliable irrigation water. The second step is to distinguish whether the observed effects of infrastructure rehabilitation on adoption were enhanced by WtM training. For this second step, we can use the subgroup analysis framework in equation (2) to determine the extent to which any observed impacts of the Infrastructure Activity on adoption of new practices were different in WtM communities.

Lastly, we plan to collect two pieces of qualitative information from WUA directors as part of the added module discussed in Chapter V with the aim of examining adoption from a different angle. The questions will ask WUA directors whether, following the improved irrigation water availability, they have observed important changes in the types of crops farmers grow and the types of irrigation practices farmers use in their WUAs. This will help triangulate an answer to the two questions posed at the beginning of this chapter and will also provide a rough measure of broader adoption rates, which we can use to assess whether the estimates for the opportunistic sample of WtM communities in the TCS are illustrative of broader trends.

# VI. REPORTING AND DOCUMENTATION

In this final chapter we summarize the timeline for implementing the evaluation and reporting the findings, plans for documenting the survey data, privacy requirements of the data and institutional review board clearance, and an overview of the evaluation team.

## A. Evaluation schedule

***Collect and Analyze Data.*** TCS data collection, including the associated survey of village mayors, was completed in May 2014. We began data cleaning and preliminary analyses of the TCS upon receiving the data file in May 2014. The final round of data collection for the WUA administrative data is underway and is expected to be finished in September 2014. We will begin analysis of the WUA data once data collection concludes.

***Final Irrigation Evaluation Reporting.*** Mathematica will prepare a final report that is submitted in three stages. First, Mathematica will provide a detailed report outline that shows the report's structure and planned contents (October 2014). After receiving MCC's comments on the outline, Mathematica will prepare a draft report (February 2015) to share with MCC and other stakeholders and discuss in detail at two stakeholder workshops: one in Washington, DC, after the draft report is complete, and another in Armenia in spring 2015, with former MCA staff, implementers, data collectors, and other interested stakeholders. In these workshops, Mathematica will present the evaluation designs and results. Following the stakeholder workshops, Mathematica will incorporate feedback and prepare the final report, which we plan to submit in April 2015. The report will include an executive summary intended for a broader audience. Once the executive summary is approved by MCC, it will be translated into Armenian as well. For the stakeholder workshop in Armenia, Mathematica will prepare a presentation in both English and Armenian.

***Economic Rates of Return.*** Updated economic rates of return will be estimated based on the findings from the irrigation evaluation. These will either be included in the evaluation report itself or submitted as a separate document at MCC's discretion, along with a spreadsheet that performs the calculations. Our estimates of the ex post economic rates of return will compare the benefits of irrigation—primarily in terms of estimated impacts on household income (as measured in the TCS)—to the costs of the rehabilitation work, using other contextual information (such as perceived sustainability of the investments) to project the future benefit streams.

We also plan to revisit the full ex ante model of benefits and costs. The difference between this approach and the approach described in the previous paragraph is that the ex ante model uses more granular input measures, such as the growth in cultivation of high-value crops, that were expected to feed into income impacts instead of using the income impact estimate itself. Looking at the benefits and costs in this different way could be useful both as a sensitivity check on the estimated economic rate of return and to explore what parameter assumptions in the ex post benefit-cost analysis model were different from the ex ante assumptions. For this analysis, we plan to involve Melik Gasparyan as a consult to update the full benefit cost models and provide an assessment after the irrigation evaluation is finalized. We anticipate that these calculations will be completed in April or May 2015.

## B. Data access, privacy, and documentation plan

***Institutional Review Board Approval.*** Health Media Lab's institutional review board reviewed and approved our data collection and analysis plans that involve human subjects, that is, the data collection plans for the TCS and the associated analyses. Approval was received on October 22, 2013 and must be renewed on October 22, 2014.

***Data Access.*** Once the irrigation evaluation is complete, we will provide the raw data, full analysis file (not for public disclosure), and public use file (with de-identified data as necessary) for the TCS along with the appropriate documentation per MCC's metadata requirements. Additionally, we will provide, at a minimum, an analysis file for the WUA administrative data that possibly cannot be made publicly available. (We are discussing with the subcontractor whether a public use file might be possible.)

***Final Programs.*** Mathematica will provide MCC with statistical program code (and documentation) for the analyses after the irrigation evaluation is complete; we anticipate that this will be submitted in June 2015. We will also provide input to MCC as needed for the development of public use files based on the FPS data.

## C. Evaluation team roles and responsibilities

Many people have helped or are helping with the data collection, analysis, and reporting for this evaluation. The team leaders for the respective tasks are as follows:

***Ken Fortson*** of Mathematica led the design of the evaluation and now leads implementation of the analysis. In this capacity he has also been closely involved in developing data collection instruments, sampling plans, monitoring progress on data collection and analysis, and communicating with MCC, MCA-Armenia, and other stakeholders.

***Ada Babloyan*** of AREG is the team lead for the TCS data collection. She has had primary responsibility for organizing and executing the TCS fieldwork, overseeing TCS data cleaning, preparing documentation, and communicating with Mathematica as needed.

***Melik Gasparyan*** of AVAG Solutions is the team lead for the WUA administrative data collection. He has primary responsibility for organizing and executing the WUA fieldwork, analogous to Dr. Babloyan's role on the TCS. Mr. Gasparyan will also be involved in the estimation of ex post economic rates of return.

***Joanne Lee*** of Mathematica contributes to the implementation of the analysis and will lead creation of the analysis files and public use files.

## REFERENCES

Bandyopadhyay, S., P. Shyamsundar, and M. Xie. 2007. "Yield Impact of Irrigation Management Transfer: Story from the Philippines." World Bank Policy Research Working Paper 4298.

Del Carpio, X. V., N. Loayza, and G. Datar. 2011. "Is Irrigation Rehabilitation Good for Poor Farmers? An Impact Evaluation of a Non-Experimental Irrigation Project in Peru." *Journal of Agricultural Economics*, vol. 62, no. 2, pp. 449-473.

Dillon, A.  2008. "Access to Irrigation and the Escape from Poverty: Evidence from Northern Mali". IFPRI discussion paper 00782.

Duflo, E. and R. Pande. 2007. "Dams." *Quarterly Journal of Economics*, vol. 122, no. 2, pp. 601-646.

Food and Agriculture Organization (FAO) of the United Nations. 2000. AQUASTAT Database on investment costs in irrigation. http://www.fao.org/nr/water/aquastat/investment/index.stm, accessed January 10, 2014.

FAO. 2009. Water Report 34. http://www.fao.org/nr/water/aquastat/countries_regions/ARM/print1.stm, accessed January 9, 2014.

Fortson, K., A. Rangarajan, R. Blair, J. Lee, and V. Gilbert. 2013. Evaluation of the Water-to-Market Activity in Armenia. Mathematica Policy Research: Oakland, CA.

Fortson, K., R. Blair, and K. Gonzalez. Forthcoming. Evaluation of an MCC- and World Bank-Financed Rural Road Rehabilitation Project in Armenia. Mathematica Policy Research: Oakland, CA.

Fortson, K., R. Blair, and A. Rangarajan. 2010. Baseline Report on the Tertiary Canal Survey. Mathematica Policy Research: Princeton, NJ.

Gulati, A, N. Minot, C. Delgado, and S. Bora. 2005. Growth in High-Value Agriculture in Asia and the Emergence of Vertical Links with Farmers. World Bank: Washington, DC.

Hodgson, S. 2007. Legislation on Water User Associations (WUA)- International Experiences. World Bank: Washington, DC.

Hussain, I., and M. A. Hanjra. 2004. "Irrigation and Poverty Alleviation: Review of the Empirical Evidence." *Irrigation and Drainage*, vol. 53, pp. 1-15.

Mukherji, A., B. Fuleki, T. Shah, D. Suhardiman, M. Giordano and P. Weligamage. 2009. Irrigation reform in Asia: A review of 108 cases of irrigation management transfer. International Water Management Institute: Colombo, Sri Lanka.

National Statistical Service. 2010. Measuring Poverty in Armenia: Methodological Explanations. National Statistical Service of Armenia: Yerevan, Armenia.

Schafer, J. L. and J. Kang. 2008. "Average Causal Effects from Nonrandomized Studies: A Practical Guide and Simulated Example." *Psychological Methods*, vol. 13, no. 4, pp. 279-313.

Van Den Berg, M., and R. Ruben. 2006. "Small-Scale irrigation and income distribution in Ethiopia." *Journal of Development Studies*, vol. 42, no. 5, pp. 868-880.

Wang, J., Z. Xu, J. Huang, and S. Rozelle. 2006. "Incentives to managers or participation of farmers in China's irrigation systems: which matters most for water savings, farmer income, and poverty?" *Agricultural Economics*, vol. 34, pp. 315-330.

Wang, J., J. Huang, L. Zhang, Q. Huang, and S. Rozelle. 2010. "Water Governance and Water Use Efficiency: The Five Principles of WUA Management and Performance in China." *Journal of the American Water Resources Association*, vol. 46, no. 4, pp. 665-685.

World Bank Impact Evaluation Group (IEG). 2008. 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.

Xie, M. 2007. Global Development of Farmer Water User Associations (WUA): Lessons from South-East Asia. World Bank: Washington, DC.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research