ICARDA

Science for resilient livelihoods in dry areas

# **Designing and Conducting Randomized Controlled Trials (RCTs) for Impact Evaluations of Agricultural Development Research:** A Case Study from ICARDA's 'Mind the Gap' Project in Tunisia

Boubaker Dhehibi[1], Jutta Werner[1], and Matin Qaim[2]
[1]International Center for Agricultural Research in the Dry Areas (ICARDA)
[2]University of Goettingen, Germany

MANUALS & GUIDELINES 1

ISBN13: 978-92-9127-504-5

**KEYWORDS:** Impact Evaluation; Development Evaluation; Program Evaluation; Agricultural Extension; Randomized Control Trials (RCTs); Tunisia; Dry Areas.

**Manuals and Guidelines**
ICARDA's Manuals & Guidelines series taps the Center's expertise to provide comprehensive advice and strategies that researchers can adopt to enhance agricultural productivity and overcome critical challenges affecting rural communities in the non-tropical dry areas.

**Suggested Citation**
Dhehibi, B., J. Werner, and M. Qaim. 2018. *Designing and Conducting Randomized Controlled Trials (RCTs) for Impact Evaluations of Agricultural Development Research: A Case Study from ICARDA's 'Mind the Gap' Project in Tunisia.* Manuals and Guidelines 1. Beirut, Lebanon: The International Center for Agricultural Research in the Dry Areas (ICARDA).

**About ICARDA**
Established in 1977, the International Center for Agricultural Research in the Dry Areas (ICARDA) is a non-profit, CGIAR Research Center that focusses on delivering innovative solutions for sustainable agricultural development in the non-tropical dry areas of the developing world.

We provide innovative, science-based solutions to improve the livelihoods and resilience of resource-poor smallholder farmers. We do this through strategic partnerships, linking research to development, and capacity development, and by taking into account gender equality and the role of youth in transforming the non-tropical dry areas.

**Address**
Dalia Building, Second Floor, Bashir El Kasser St, Verdun, Beirut, Lebanon 1108-2010.
www.icarda.org

# Contents

# 1. Introduction

Experience suggests that smallholder farmers are unlikely to adopt new innovations without improved models of technology transfer. However, the question of how to design innovative and cost-effective technology transfer strategies has not yet been sufficiently addressed. There is a risk that if we do not critically analyze and carefully manage the implementation of current approaches we will simply repeat past mistakes and fail to strengthen the resilience of farmers, enhance climate change adaptation/mitigation and maintain or raise yields.

An ICARDA initiative—'Mind the Gap'—is testing the delivery of innovative technology packages to rural communities in Tunisia using a Randomized Controlled Trial (RCT) approach. RCTs are often called the 'gold standard' of evaluation methods, as it is the only evaluation method that allows a comparison of outcomes **with and without** a particular intervention, while avoiding selection bias due to observed or unobserved factors.

Like all impact evaluations, the main purpose of an RCT is to determine whether a program has an impact, and more specifically, to quantify how large the impact of the intervention is. Impact evaluations typically measure program effectiveness by comparing outcomes for those who received the program (individuals, communities, schools, etc.) against those who did not.

This document is designed to help researchers apply RCTs so they can gain a more accurate insight into the impacts of different extension strategies in different locations. It provides information on the benefits of an RCT approach in comparison to other impact evaluation models; provides a step-by-step implementation guide and a framework to avoid challenges; and demonstrates how an RCT approach was implemented within the context of the 'Mind the Gap' initiative.

The '**Mind the Gap**' initiative aims to improve dissemination strategies to increase technology adoption by smallholder farmers. The initiative compares different extension approaches and evaluates their impacts on technology adoption rates and household livelihoods. The project targets training in the following areas: agriculture, business development, and organizational development. It also tailors business development training to women.

# 2. What is a Randomized Controlled Trial (RCT)?

A randomized evaluation is a type of impact evaluation that uses random assignment as part of a study's design. The randomization limits selection bias. In an RCT a program or policy is viewed as an intervention in which a treatment—the elements of the program/policy being evaluated—is tested to see how well it achieves its planned objectives, as measured by a previously predetermined set of indicators. The different comparison groups of the experimental design—similar to an agricultural experimental field design—allow researchers to determine any impacts of the treatment when compared with the no treatment (control) group, while other variables remain constant.

Like all impact evaluations, the main purpose of RCTs is to determine whether a program has an impact, and more specifically, to quantify how large the impact of the intervention is. Impact evaluations typically measure program effectiveness by comparing outcomes for those groups who received the program (individuals, communities, schools, etc.) against those who did not. RCTs have a longer history in the medical sciences, but have recently also become popular in development evaluations and the agricultural development context (Ashraf et al. 2009; Duflo et al. 2011; Saenger et al. 2014). There is a long history of using RCTs to inform public policy decisions, especially in health, but it is only recently that they have been used more widely in development evaluations. The randomized experiments are also known as (1) Random assignment studies, (2) Randomized field trials, (3) Social experiments, and (4) Randomized controlled experiments.

**Why is Randomization so Important in RCTs?**
The purpose of randomizing treatments among experimental units is (Fisher 1935):

1. To guard against any use of judgement or systematic arrangements leading to one treatment being disadvantaged (i.e., to avoid bias).
2. To provide a basis for the standard methods of statistical analysis, such as significance tests.

Generally, when implementing development activities, research organizations do not randomize, but instead choose participating farmers based on pre-selected criteria (e.g. members of a cooperative), or they let farmers themselves decide whether they want to participate or not. However, in both cases participants ('treated') will differ systematically from non-participants ('control'). As a consequence, when we find differences in outcomes (such as yield) between treated and control groups, we do not know whether these are due to the treatment or to pre-existing systematic differences. This can lead to selection bias in the impact evaluation. For instance, a positive impact is detected, but maybe the farmers selected to participate were more progressive farmers and had already achieved higher yields before the treatment, compared to non-participating farmers. Such selection bias can be avoided through randomization. RCTs provide a response to the question of causality since they help to clarify whether or not observed differences in yield are really due to development interventions.

A very simple RCT assigns some subjects to one or more treatment groups, and others to a control group. The treatment group participates in the program being evaluated and the control group does not. After the treatment group experiences the intervention, an RCT compares effects on the two groups (treatment vs control) by measuring the difference, and assesses the impact between the two groups on the outcome of interest. This difference is subsequently considered an estimate of the program's impact.

**Random Sampling vs Random Assignment**

- Random assignment should not be confused with random sampling.
- Random sampling refers to how a sample is drawn from one or more populations, whereas random assignment refers to how individuals or groups are assigned to either a treatment group or a control group.
- RCTs typically use both random sampling (since they are usually aiming to make inferences about a larger population) and random assignment (an essential characteristic of an RCT).

# 3. When is it Appropriate to use RCTs?

While RCTs are considered a rigorous way of measuring impact in certain circumstances, they are only one part of a wider array of evaluation tools. In the section below, we describe the relevant conditions for using the RCT approach in impact evaluations.

**RCTs should be planned from the beginning of the program:** An RCT needs to be planned from the beginning of a program's implementation, and participation in the program needs to be carefully controlled and monitored with the experiment in mind. RCTs cannot be undertaken retrospectively.

**RCTs need a large sample size:** An RCT can only be used when the sample size is big enough to capture the impacts of the program with sufficient precision, and the study design must have sufficient statistical power. Given that power increases with a larger sample, one critical part of the process of designing an RCT is to perform power calculations, which indicate the sample size required to detect the impact of the program. It is important to have enough observations in the different treatments and control groups to prove effects that are statistically significant.

However, while larger sample sizes are always preferable from a statistical point of view, it is also important for researchers to assess realistically how many farmers can be included, given project constraints in terms of financial and human capacity. For this reason, the sample size will be a compromise between feasibility and statistical power.

**Relevant Conditions for Using RCTs in Impact Evaluations**

- RCTs should be planned from the beginning of the program.
- RCTs need a large sample size.
- RCTs should be undertaken following formative research or evaluations to guide the entire process.
- RCTs must be appropriate to the nature of the program being assessed.
- RCTs need a very close follow up and monitoring, as they are intensive in time requirements and implementation costs.

**RCTs should be undertaken following formative research or evaluations:** Using an RCT framework to evaluate and quantify the impact of an immature program is likely to be inappropriate and, under most circumstances, an RCT should not take place until the program design has been adequately developed.

**RCTs must be appropriate to the nature of the program being assessed:** RCTs are best used for programs that seek to achieve clear, measurable impacts that can be attributed to a specific intervention, or set of interventions, and which lend themselves to causal pathway analysis. RCTs are not well suited to emergent programs, and are not able to achieve clear, common and quantifiable results that are hard to measure.

# 4. How to Conduct a Randomized Controlled Trial

The RCT method can be implemented, and impacts estimated, using the following steps:

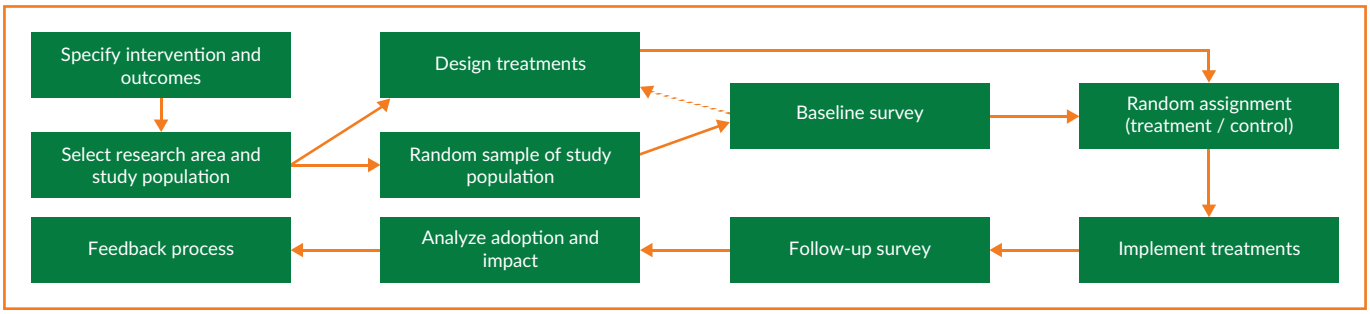


*Figure 1: Overview of conducting an RCT*

**Step 1. Specify interventions and outcomes:** As with any impact evaluation, an RCT should start by clearly specifying what is being evaluated and why, and the outcomes and impacts based on RCTs have to be well planned before the treatments are run. An extremely close coordination between researchers and implementers of the treatments is required.

**Step 2. Select research area and study population, and design the study carefully:** It is important when designing an RCT that the eligible population and the unit of assignment for randomization purposes are clearly identified, and that consistency is ensured.

**Step 3. Design treatments and randomly assign people to treatment or control:** There are two steps of randomization: the random sample of the study population and the random assignment of the treatments (see step 5).

**Step 4. Conduct the baseline survey:** Collect baseline data from treatment as well as control groups. Before or after the random assignment of participants, evaluators usually conduct a baseline survey to generate the data that will be used as the basis for the follow-up survey after the treatments, and perhaps subsequent comparisons.

**Step 5. Random assignment of treatments and control:** The key to randomized experimental research design is in the random assignment of treatments and control. This process will assure the similarity between groups when treatment begins.

**Step 6. Implement treatments and monitor processes closely so that the integrity of the experiment is not compromised:** For a very efficient use of RCTs a rigorous monitoring system for the program's implementation is needed. All people within one treatment have to receive the same approach. A proficient monitoring and follow-up system, by using collected data on the implemented intervention in the treatment group, will help to make sure that it is always evident for the research who has been assigned to which treatment and received which approach.

**Step 7. Collect data about implementation (and possibly mid-term outcome data) and prepare midterm reports:** Data that provides information on implementation should be collected, possibly through a mid-term survey, which will usually focus on process aspects. Such a survey may also be used to provide initial estimates of program impacts (through this monitoring and evaluation process) if it is not premature to do so.

**Step 8. Collect data on impacts:** Following the implementation of the program, a follow-up survey should be conducted. The timing of the follow-up survey will depend on the theory of change, as to how long it will take for the expected impacts to occur.

**Step 9. Data Analysis:** Analyze data, and report the intervention's effects on adoption and impact, with a special emphasis on the statistical significance and the magnitude of each effect.

**Step 10. Feedback process:** Develop a strategy on how the results of the RCT can be implemented by relevant stakeholders and decision makers and share it with them.

**Overcoming Challenges**
Constraints, challenges and information on what RCTs cannot do in impact evaluations is summarized in the following section:

- **Avoiding contamination:** The implementation of the experimental design is a challenging procedure, and contamination between the different treatments and the control group has to be avoided since this would severely impact the results/effects. To avoid contamination a spatial separation should be applied as much as possible. Avoiding the inclusion of several treatment groups within the same village can also be achieved through cluster sampling.
- **Restricting number of options to be tested:** Because of the challenging experimental design, the number of different options that are going to be tested is limited. Development practitioners tend to opt for testing a large amount of different options, which is not feasible under practical implementation conditions. In order to be comparable, the implementation of the treatments should be as similar as possible. The implementing extension service/agency/NGO should for this reason be equipped with an exact manual/protocol they can follow to make sure that the message given to farmers/households is consistent.
- **Maintaining the integrity of the design:** Even if random assignment is put in place, there are several potential challenges, such as: (1) low take-up of the intervention, (2) lack of compliance with intended procedures, (3) contamination of the control group by other interventions affecting similar outcomes or through self-contamination, and (4) change in the design or location of the program being evaluated. Most of these problems can be dealt with at the analysis stage, but the evaluators need to collect the necessary data in order to be aware of the issues and be able to address them in an appropriate manner.
- **Failing to adjust standard errors when a cluster design has been used:** This is a common technical error, which artificially increases power and may incorrectly conclude that an intervention is working when in fact it is not.
- **Excessive focus on the average treatment effect:** An RCT provides an unbiased estimate of the mean effect of an intervention. This is, however, rarely the finding of most interest to policymakers, who are often particularly interested in how effective an intervention is for particular subgroups, especially those interventions that address equity issues.

- **Opposition to random assignment:** There is often opposition to random assignment from the implementing staff. Having agreement from management may be insufficient to gain the cooperation of fieldworkers.

# 5. Analysis and Presentation of Results

In order to secure the quality of empirical findings, the use of the RCT method requires quality assurance. Indeed, there is a strong need to provide an accurate methodological framework, and clear and understandable findings. Another important aspect is to focus on the detailed description of the intervention being evaluated rather than focusing on the RCT method. The evaluation process generates relevant information from this diagnosis analysis. Thus, it is worth indicating that a detailed description of the intervention allows for the Theory of Change (ToCh) to be linked with an analysis of the results. When reporting the findings of an RCT, therefore, a detailed description of the theory of change should also be provided. Furthermore, it is recommended that when detailing the methodology the sampling is described, as well as the random allocation method. In this description it is important that the report describes both the number of clusters and the number of households and/or individuals in the treatment and control groups.

The impact estimates can be reported using Difference-in-Differences (DiD) analysis (a practical tool to estimate treatment effects which compares differences between treatment and control groups). The findings from the DiD analysis can be applied for the entire sample, as well as for subgroups, in order to analyze heterogeneous effects. It is also recommended that such findings need to be linked to the ToCh. This is a way to explore if the applied analysis supports the ToCH, or not. If such an analysis does not support the ToCh, it is necessary to understand which assumption behind the ToCh was not validated, and to indicate the possible reasons—from both within and outside the ToCh—that could have generated the results. (DiD analysis requires data from baseline and follow-up surveys and is preferable when such data are available. If baseline data were not collected and the randomization process was properly implemented, meaningful impact analysis can also be carried out with the follow-up data alone).

The final step is to identify concrete policy-relevant implications and recommendations which should explicitly be linked to the empirical findings and data analysis, and to include further discussions on whether the results can be generalized to different/similar locations (similar social agro-ecological zones).

# 6. Case Study: RCTs and Real World Impact—the ‘Mind the Gap’ Project in Tunisia

**Context**
Using an RCT within the framework of the 'Mind the Gap' project provides evidence-based advice to the Tunisian Government and development cooperation stakeholders in order to improve their current technology transfer strategies. The project aims to improve agricultural extension to promote technology adoption and enhance the welfare of farm households.

**Hypothesis**
We are not primarily interested in whether or not extension as such has a positive impact; rather we want to compare different extension approaches in order to establish which approach has the most significant impact.

**Research Objectives**
- Generate knowledge about how to improve extension in Tunisia and beyond.
- Assess the impact of extension approaches on:
  - Technology adoption.
  - Agricultural productivity (output, production cost).
  - Household welfare/livelihoods (income, nutrition, gender).

**Randomization Process**
The project adopts two steps of randomization:

1. Random sample of the study population.
2. Random assignment of the treatments: evaluations based on RCTs have to be well planned **before** the treatments are run, and this requires extremely close coordination between researchers and implementers of the treatments.

**Conceptual and Methodological Framework**
For more than two decades, the dominant questions related to agricultural extension focused on organizational and financial aspects of agricultural extension, or cross-cutting issues like gender participation or empowerment. There is also some research that has analyzed the impacts of specific extension approaches in particular contexts. While such analysis helps to evaluate whether or not a specific approach was successful in a particular context, it does not include how the approach could be further improved or how alternative approaches would have performed in the same situation. As a result, the research question of how to design more cost-effective technology transfer strategies has not yet been sufficiently addressed. Given past experiences, widespread adoption of new technologies among smallholders will not happen without improved models of technology transfer. This is especially true for system technologies that are often quite knowledge-intensive. Unless we critically analyze and carefully manage the implementation of approaches that are presently being favored, we will simply repeat mistakes from the past at the expense of farming communities.

Past impact studies have mostly used observational data, trying to control for possible selection bias through statistical techniques. As mentioned above, we use experimental approaches by designing and implementing an RCT. This has two advantages: First, the randomization allows researchers to attribute impacts to particular treatments in a precise way. Second, unlike the situation with observational data, where only impacts of existing interventions can be evaluated, RCTs allow the evaluation and comparison of a set of alternative interventions (options) which is extremely useful to further improve technology transfer models. So far, in the Tunisian case, RCTs have not been used to evaluate and compare agricultural extension and technology delivery.

The 'Mind the Gap' research project is testing new and existing models of transferring innovative technology packages to smallholder farmers by using an RCT approach, aiming to improve dissemination strategies to increase technology adoption by smallholders. Understanding which extension approaches have the greatest success rate help to improve future agricultural technology dissemination efforts. The 'Mind the Gap' project compares different extension approaches and evaluates their impacts on technology adoption rates and farm household livelihoods. In addition to agricultural training, extension approaches in this Project include business training for farmers, training for organizational development and business training tailored to women. A policy manual for extension services is also being developed. Scaling up to other regions will be facilitated with interactive similarity maps that identify similar socio-cultural and environmental contexts across the West Asia and North Africa (WANA) region.

**'Mind the Gap' Project: The Challenge**
Proven agricultural technologies that can improve lives often have low adoption rates due to lack of effective service delivery systems and enabling policy environments. Agricultural extension is a common method to introduce technologies. However, little is known about which extension approaches are the most effective with farmers.

To improve the rigor of the comparison, the project is implementing an RCT approach. RCTs compare randomly selected groups that receive the extension approaches being tested ('treatment group') with those that do not receive the approaches ('control group'). Different treatment groups receive different types of extension approaches. These groups are compared with each other and against the control group receiving no treatment. The technology transfer models that we test comprise three components: access to technology packages and information related to the technology package, access to capacity building (entrepreneurial training and organizational development), and entrepreneurial training/information related to the technology package targeting women. These three components are combined in various ways, and the combinations are implemented in different treatment groups to test and compare their individual and combined effects.

Four combinations of treatments of the technological transfer models were developed with careful consultation with regional scientists using observations and results across semi-arid Tunisia from past and existing research projects. Expert-based estimations and assessments demonstrated the appropriateness of 140 treatment units (farmers) for each treatment in the governorates of Zaghouan and Kairouan. This group size provides adequate statistical power for inferring the difference in outcome indicators driven by the treatments implemented. The masking effects of variation in non-experimental factors on the RCT results can be minimized with the aid of environmental similarity mapping. The mapping has delineated relatively uniform spatial clusters of socio-ecological and cultural conditions.

In total, we compare four different treatments with and without certain components included, and one control without any treatment. Each treatment is implemented with 140 randomly selected farmers in the two target Governorates. Together with the control group, which also consists of 140 randomly selected farmers, the total sample size is 700 farm households. The different treatments and their individual components are rigorously evaluated in terms of their costs and effects on innovation adoption, farm productivity and household livelihoods (especially income and gender roles). This approach is unique in that it allows for:

- Randomization that allows impacts to be attributed to a particular treatment.
- One or different treatment groups and a control group.
- Comparison and evaluation of different extension approaches.

### Different treatments for the five groups of farmers

| Groups | 1 Control | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **Number of farmers** | 140 | 140 | 140 | 140 | 140 |
| **Component 1** | Ø | barley & feed blocks ♀♂ | barley & feed blocks ♀♂ | barley & feed blocks ♀♂ | barley & feed blocks ♀♂ |
| **Component 2** | Ø | Ø | entrepreneurial & organizational training ♀♂ | entrepreneurial & organizational training ♀♂ | Ø |
| **Component 3** | Ø | Ø | Ø | ♀ | ♀ |

Note: ♀: Approach for rural women; ♀♂: Approach for women and men.

**Treatment groups:**

- **1st group:** Control.
- **2nd group:** Access to technology packages (barley & feed blocks) and technical training related to the two technologies.
- **3rd group:** Access to technology packages (barley & feed blocks) and technical training related to the two technologies + access to entrepreneurial and organizational training.
- **4th group:** Access to technology packages (barley & feed blocks) and technical training related to the two technologies + access to entrepreneurial and organizational training + access to information related to the technical package targeting women.
- **5th group:** Access to technology packages (barley & feed blocks) and technical training related to the two technologies + entrepreneurial training and information related to the technical package targeting women.

Data for the impact analysis are collected through a baseline survey (before implementation of the treatments) and a follow-up survey (after implementation). Significant differences in technology uptake between women and men will be evaluated to identify the most successful technology transfer model to empower women. A more detailed explanation on the different treatments for the five groups of farmers are presented in the above table.

Ethical aspects of the RCT experiment are cautiously considered. The control group will receive project benefits with a time lag of one and a half years, as this group does not receive any benefit from the treatment package but is supposed to participate in the baseline and follow-up surveys. This time lag corresponds to an accepted practice in development projects. Such projects start in one community and scale the measures gradually to other communities. Participating farmers were selected randomly through a computer-based drawing to improve transparency. All personal data collected will remain anonymous at all times.

# 7. Further Reading

Ashraf, N., X. Giné, and D. Karlan. 2009. "Finding missing markets (and a disturbing epilogue): evidence from an export crop adoption and marketing intervention in Kenya." American Journal of Agricultural Economics 91: 973-990.

Deaton, A. 2010. "Instruments, Randomization, and Learning about Development." Journal of Economic Literature 48: 424-455.

Duflo, E., M. Kremer, and J. Robinson. 2011. "Nudging farmers to use fertilizer: theory and experimental evidence from Kenya." American Economic Review 101: 2350-2390.

Duflo, E., R. Glennerster, and M. Kremer. 2007. "Using Randomization in Development Economics Research: A Toolkit." Handbook of Development Economics 4: 3895-3962.

Duflo, E., R. Glennerster, and M. Kremer. 2006. Using Randomization in Development Economics Research: A Toolkit. J-LAB. See: http://www.povertyactionlab.com/papers/Using%20Randomization%20in%20Development%20Economics.pdf

European Evaluation Society. 2007. EES Statement: The importance of a methodologically diverse approach to impact evaluation – specifically with respect to development aid and in development interventions. Prague. See: http://www.europeanevaluation.org/news?newsId=1969406

Evergreen, S. (n.d.). Visualise Data. Better Evaluation. Retrieved May 2017 from http://betterevaluation.org/plan/describe/visualise_data; http://www.betterevaluation.org/en/plan/approach/rct

Gertler, P. J., S. Martinez, P. Premand, L. B. Rawlings, and C.M.J. Veermeersh. 2011. Impact Evaluation in Practice. Washington, DC: World Bank. See: https://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf

Leeuw, F. and J. Vassen. 2010. Impact evaluations and development. Nonie guidance on impact evaluation. Washington DC: Independent Evaluation Group, World Bank. See: http://www.worldbank.org/ieg/nonie/

Mind the Gap. 2016. Improving Dissemination Strategies to Increase Technology Adoption by Smallholders. Project Brief. Beirut: International Center for Agricultural Research in the Dry Areas (ICARDA). See: https://apps.icarda.org/wsInternet/wsInternet.asmx/DownloadFileToLocal?filePath=Project_summaries/ICARDA-Mind_the_Gap-flyer.pdf&fileName=ICARDA-Mind_the_Gap-flyer.pdf

Saenger, C., M. Torero, and M. Qaim. 2014. "Impact of third-party contract enforcement in agricultural markets – a field experiment in Vietnam." American Journal of Agricultural Economics 96, 1220-1238.

Urbaniak, G. C., and P. Scott. 2013. Research Randomizer (Version 4.0) [Computer software]. Retrieved from http://www.randomizer.org/

White, H. 2013. "An introduction to the use of randomized control trials to evaluate development interventions." Journal of Development Effectiveness 5(1), 2013: 30–49. See: http://www.tandfonline.com/doi/pdf/10.1080/19439342.2013.764652

ICARDA

Science for resilient livelihoods in dry areas

AVFA

IRESA

Institution de la Recherche et de l'Enseignement Supérieur Agricoles

INRAT

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

With the financial support of

Federal Ministry for Economic Cooperation and Development