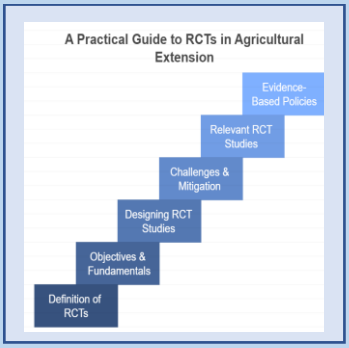
An-Najah University Journal for Research – A
Natural Sciences

Enhancing Agricultural Extension Programs: A Practical Guide to Designing and Implementing Randomized Controlled Trials

Amin ABU-ALSOUD[1,*], Ameur Mehrez[2], Houcine Bichni[3]

Accepted Manuscript, In press

Abstract: Evidence-based practices in agricultural extension programs must be guided by precise evaluation strategies and reliable data. Methods like randomized controlled trials (RCTs) are vital for assessing specific program interventions. However, not all studies use experimental designs, and their quality varies, as noted in a 2010 World Bank study. RCTs, known for providing robust impact assessments, involve complex steps and are applied in various fields. This guide focuses specifically on RCTs used for agricultural extensions; a topic covered less often than clinical trials. More precisely, this guide provides (1) definitions and key steps for RCTs, (2) a simplified RCTs study design methodology, (3) an exploration of challenges and mitigation, and (4) a summary of relevant RCTs studies in extensions. This guide is formulated to enhance the present understanding of RCTs and elaborate on the ways to use them in their application in agricultural research. Furthermore, the goal here is to make sure that our policies and methods are presented with solid evidence. This approach is meant to encourage the implementation of agricultural practices that are not only more effective but also sustainable.



Keywords: Randomized Controlled Trial, Impact Evaluation, Agricultural Extension, Randomization, Research Methods

Introduction

Studying and gaining a complete understanding of RCTs (Randomized Controlled Trials) in agriculture helps us really understand how they are used and what they can do in farming and making policies to improve agricultural outcomes. These trials are famous for their careful way of collecting data, and they have had a huge impact on policies in various fields like social sciences and development studies. Additionally, RCTs hold even greater significance in medical research, where they are considered the gold standard for evaluating interventions and treatments. However, in the context of agricultural research, their adoption is increasingly recognized as a crucial tool for generating robust, evidence-based policies and practices. This got a lot of attention worldwide when Abhijit Banerjee, Michael Kremer, and Esther Duflo won the Nobel Prize in Economics in the year 2019 for using RCTs to overcome poverty issues. [1] (p. 253).

In agricultural research, RCTs are highly valuable and have been widely used to evaluate the impact of various technologies and interventions, such as different types of fertilizers, pesticides, and animal feeds. This allows for precise measurement of intervention effectiveness at the field level [1] (p. 253). Despite the proven impact of RCTs, evidence suggests that they are underutilized in agricultural policymaking, particularly in the Middle East. According to [2] (p 351), experimental design has gained increasing importance in impact evaluation, yet its application remains limited in agricultural advisory services. This underuse highlights the potential for agricultural economists and researchers to leverage RCTs more extensively to inform and develop resilient, sustainable agricultural policies [1] (p. 253).

Agricultural extension programs are of great importance in agricultural development. This is evident through the dissemination of knowledge and best practices, which results in improved farmer behaviors, productivity, and livelihoods. These programs are a binder in sustainable agricultural development by merging practical research to field application [3-10].

Nevertheless, the question of whether international aid contributes to the success of agricultural development programs or not has been raised through suboptimal outcomes that led to the call for the need for impact evaluations. Evaluations of this type are common in the social sector, but their usage in agricultural extension is limited, highlighting the need for more advanced knowledge and methodological development [11].

RCTs, famed for their solid approaches to establishing causality, are essential in building evidence-based programs and policies in agricultural extension research. A 2010 World Bank review highlighted a significant gap in this area: from almost 25 thousand studies, only 86 were revealed to be true impact evaluations, however, only 6% of them were experimental. This underscores the necessity to broaden the implementation of experimental methods in agricultural extension research [12].

1 Higher School of Agriculture of Kef (ESAK), University of Jendouba, Tunisia.
*Corresponding author: abualsoudamin@gmail.com
2 mameur2000@yahoo.es
3 National Institute for Agricultural Research of Tunisia (INRAT), University of Carthage, Tunisia. houcine.bchini@gmail.com

The RCTs design that was meant to compare different extension methods has not been applied due to high costs, as the research budget must cover data collection, evaluation, and the actual implementation of extension treatments [13] (pp. 100–115). In countries with limited capacities, pilot studies using experimental methods are vital for policymakers, especially in developing nations where resources are scarce. Rigorous pilot RCTs studies could prevent policies from underperforming and optimize resource utilization [14].

This article rigorously examines and synthesizes insights from existing research on RCTs in agricultural extension program evaluation, integrating scholarly works from [1, 11, 15, 16]. The objective is to facilitate the design and implementation of effective RCTs in this sector, providing a detailed, accessible guide to enhance future scholarly work in crucial areas of agricultural development.

Objective

We present concise and focused key information, as well as practical guidelines and tips that can help agricultural extension researchers plan and conduct RCT-based studies. This guide provides [1] definitions and key steps for planning and conducting RCTs, [2] simplified methodologies for RCTs study designs, [3] an overview of potential challenges and strategies for mitigation, and [4] a summary of relevant RCTs studies in extensions. This guide caters to a diverse academic audience, encompassing both newcomers and seasoned researchers in the field of agricultural extension. In this guide, more materials are added to provide practical examples and detailed insights that are helpful for even those who are not familiar with these practices.

Distinct value of this guide

Building on the foundational work of Gertler et al. [17] and Duflo, Glennerster, and Kremer [18], our guide stands out by extending RCTs methodologies to the complex, often-overlooked challenges of evaluating agricultural extension programs. Unlike previous research, which primarily focused on product-based interventions, our guide addresses "soft" components like knowledge dissemination, decision-making, and long-term behavioral change—key factors in agricultural success but difficult to measure with traditional RCTs frameworks.

What truly sets this guide apart is its focus on bridging academic research with practical application. With step-by-step instructions and a strong emphasis on capacity building, it makes advanced RCTs designs—such as multi-arm trials and cluster randomization—accessible to field practitioners. Even those with limited technical expertise can apply these methodologies effectively in agricultural settings.

Additionally, our guide provides detailed guidance on power and sample size calculations, randomization methods, and a comparative analysis of impact evaluation methodologies. It positions RCTs within the broader context of agricultural research, showing how they complement other experimental designs.

To ensure practical application, we include selected software tools for key calculations and examples to clarify complex concepts. We also address logistical barriers, farmer engagement, and variability in field conditions, providing real-world solutions and mitigation strategies for consistent implementation.

By integrating case studies and offering a comprehensive framework for assessing both short- and long-term impacts, our guide serves as a valuable tool for those seeking to measure the lasting effects of agricultural interventions, particularly in regions facing budget constraints and efficiency concerns.

RCT Fundamentals

Definition of RCTs

RCTs, short for Randomized Controlled Trials, are a quantitative research methodology that uses controlled conditions and random allocation to reveal or affirm causal relationships in studies, offering the utmost precision and consistency. These trials involve grouping cohorts (study participants or experimental units) into intervention and control groups, thereby broadly assessing the effectiveness of agricultural extension services [19] (p. 1). With the advancement in agricultural development, RCTs have been emerging as trustworthy tools. These trials involve grouping cohorts into intervention and control clusters, hence broadly assessing the effectiveness of agricultural extension services [20] (p. 1). RCTs are specifically designed to minimize selection bias and accurately measure the impact of interventions. They achieve this by comparing treatment and control groups, where random assignment ensures comparability across these groups, thereby eliminating any potential bias. This approach is crucial for deriving clear, unbiased insights from a study, ensuring that the results are solely attributable to the intervention being tested [15] (p. 4), [1] (p. 254).

In cluster randomized controlled trials (c-RCTs), randomization is not conducted on individual patients [farmers] but rather on groups of patients [farmers], such as those defined by geographical area, hospital, or school [21] (p. 132).

Evolution of Randomized Controlled Trials

The origins of randomized evaluations can be traced to the 18th century when James Lind's experiment with sailors revealed that citrus fruits can prevent scurvy, enabling modern clinical trials [22]. The perception of treatment and control in experiments was further developed in the 1920s by Neyman and Fisher, whose work in the field of agriculture allowed Fisher's influential publication, "The Design of Experiments" [23], with contributions by Laura Feeney and Claire Walsh. The use of government-sponsored randomized trials expanded in the latter half of the 20th century, shifting the focus of research subjects from plants and animals to human participants [23]. This evolution marked the beginning of an era in which randomized evaluations have become a cornerstone methodology in various fields that are continuously adapting and growing over time [1] (p. 254).

What Can RCTs Measure?

RCTs generally address one evaluation question of interest—effect size—while potentially neglecting other important evaluation questions [11] (p. 3).

Benefits of Employing RCTs:

The primary advantage of RCTs lies in the process of randomization itself, which significantly mitigates confounding factors, both known and unknown. While non-randomized studies can adjust for identifiable confounders, dealing with unmeasured or unknown confounders is considerably more challenging, though certain methods exist for this purpose. Randomization enables causal inferences to be drawn about the

relationship between the intervention or exposure and the observed outcome.

The additional advantages of RCTs arise from the controlled and prospective design. Their structure makes it possible to fine tune the different parameters of the intervention, such as its schedule, frequency, and length. What's more is that blinding, the concealment of the allocation of treatments, is often possible in RCTs, thus reducing bias and ultimately increasing the reliability of the results.

Moreover, statistical reliability and causality are two areas where RCTs are highly regarded, as noted by Athey and Imbens [24]. They excel at providing credible causal inferences, which is essential for reducing biases that may distort observational studies. Bhide et al. [19] underscored the effectiveness of RCTs in this regard and emphasized their superior credibility compared to observational approaches.

Dhehibi et al. [15] acknowledge that RCTs have a strong statistical foundation and significantly advance evidence-based policymaking. Furthermore, Bhide et al. [19] highlight that the replicability and generalizability of RCTs results are important advantages that enhance their applicability.

Furthermore, Groenwold et al. [25] emphasize that ethical elements inherent in RCTs, such as fair random allocation, guarantee the integrity of studies and assist decision-makers in accurately identifying the true impacts of interventions. Guo et al. [26].provide additional evidence that RCTs are less prone to bias than observational research, reinforcing the superiority of randomized experiments in obtaining reliable results.

Downsides to Applying RCTs

RCTs are limited in scope and are likely to ignore some crucial aspects of a program evaluation. By way of illustration, they do not comprehensively analyze internal efficiencies of the activities, do not explain the mechanism by which the programs produce results, and not sufficiently assess how well intervention activities match the core needs and objectives of the beneficiaries. Besides these limitations, there are a lot of other essential evaluating questions that RCTs usually don't include, which indicates the need for a more comprehensive approach that would cover all the necessary aspects [11] (p. 3).

Doss [27] highlights the limitations of non-experimental approaches, which are primarily their descriptive nature and inability to fully tackle complex research questions. In contrast, Guo et al. [26] focus on filling gaps in the literature by considering factors such as gender and secondary employment. They emphasize that the methodology of RCTs offers major advantages, including the prevention of selection bias, which leads to more accurate evaluation of treatment effects [26].

Characteristics of RCTs

RCTs have particular characteristics which make them different from the rest. First of all, they would possibly have the strict inclusion and exclusion criteria which would lead to the participant pool that is not completely representative of the general population, thus, limiting the generalizability of study's outcome. In addition, the RCTs are mostly carried out in the controlled, perfect environments that are very far from the real world conditions, thus limiting the feasibility of their results [28] (p. 582).

Furthermore, because of the complex planning and execution of RCTs, they are costly and time-consuming. In certain situations, such as rare conditions or when randomization is impractical or ethically dubious, conducting RCTs is not feasible. These aspects collectively challenge the effectiveness and relevance of RCTs in some scenarios [28] (p. 582).

Additionally, RCTs require large sample sizes to ensure their precision and sufficient statistical power. Conducting power calculations is crucial to determining the sample size needed to identify statistically significant effects in both treatment and control groups [15] (p. 6).

RCTs Stand Out from Other Impact Evaluation Methods

In impact evaluations, experimental methods like RCTs are preferred due to their minimal, testable assumptions and ability to yield unbiased, precise estimates with an adequate sample size. Propensity score matching, regression discontinuity design, and difference-in-differences are examples of quasi-experimental techniques, are more assumption-heavy and challenging to validate [1]. RCTs are widely regarded as the gold standard in this field, as noted by [17]. For more information, consult the comparative analysis found in Appendix A, Table A1.

Comparison of Experimental Designs:

This section provides an overview of key experimental designs, highlighting their core principles and demonstrating how they complement RCTs in agricultural extension evaluations. By exploring the use of Completely Randomized Design (CRD) and Randomized Complete Block Design (RCBD), we illustrate how these designs can enhance the rigor and reliability of RCTs outcomes in diverse agricultural contexts.

Definition of concepts and key features:

RCTs

RCTs are a type of experimental design often used in clinical and healthcare research.

It involves randomly assigning participants into either a treatment group or a control group, which helps eliminate bias and ensure that the differences observed are due to the treatment rather than external factors.

Key feature: Random assignment of individuals to different groups (treatment or control) to evaluate the effectiveness of interventions.

CRD

CRD is an experimental design used in fields like agriculture, psychology, or laboratory experiments where subjects or units are randomly assigned to treatments [29].

In CRD, all experimental units are assigned randomly, without any restrictions, to various treatment groups.

Key feature: Random assignment of units to treatments, but it does not account for any other sources of variability, such as environmental or blocking factors.

RCBD

RCBD is a type of experimental design often used in fields like agriculture or engineering.

In RCBD, subjects (or experimental units) are grouped into blocks that are as homogeneous as possible. Then, within each block, subjects are randomly assigned to treatment groups. This helps control for variability within blocks, leading to more accurate results.[30]

Key feature: Blocking accounts for known variability among subjects, and randomization happens within each block.

How These Designs Relate to RCTs:

RCTs generally refer to a broad concept of randomized trials but do not dictate the specific design structure. They can use CRD or RCBD or any other design as part of the trial design depending on the experimental context:

RCTs with CRD:

If RCTs are performed using a CRD, participants or experimental units are randomly assigned to treatment or control groups without considering any blocking or stratification [29]

This is commonly used when there are no significant covariates or subgroups that need to be accounted for.

RCTs with RCBD:

In agricultural experiments, RCTs can be conducted using a RCBD to account for factors that may introduce variability, such as location, soil fertility, or plant age. In this design, experiment units (e.g., plots or plants) are grouped into blocks based on these characteristics. Within, each block, random assignment is used to allocate units to treatment or control groups.

How These Designs Relate to RCTs in Agricultural Extension Evaluation:

RCTs with CRD: In a village-level program evaluating the impact of a new agricultural training method, entire villages could be randomly assigned to receive either the new training (treatment) or no training (control), without considering differences between villages. This would be an RCTs with a CRD design.

RCTs with RCBD: If the program evaluators believe that villages located in different geographic regions might experience different outcomes based on access to water or land fertility, they might first divide the villages into blocks based on region. Within each block, villages would be randomly assigned to receive the new training or not. This would be an RCTs with an RCBD design, controlling for variability in geographic location.

Therefore, RCTs in agricultural settings are not limited to simple control-intervention comparisons but can be structured using various designs to address complex research questions and improve the validity of the results. For detailed examples illustrating how CRD and RCBD are applied within RCTs, please refer to Appendix B.

Sequential Steps in RCTs Design and Implementation

Figure 1 presents a detailed framework for executing RCTs within agricultural extension research, outlining a systematic approach for conducting experimental studies. This structured process commences with a clear definition of the intervention being tested, along with a detailed articulation of the anticipated outcomes. This first step is essential to provide a clear focus and seek the trial's later stages.

After determining what the intervention will be, we proceed to create it. Therefore, planning entails the development of the separate parts of the intervention, such as training programs, material provision, or the introduction of new farming methods. It's all about considering that these should be carefully tailored to our specific requirements.

The next step is the selection of a topic and a target group. It is crucial to accomplish this goal so that the readers can easily determine the spot on the map and effectively make up the main group of the target audience. Generally, such a choice involves a number of factors such as agricultural practices, socioeconomic characteristics, and environmental conditions that are common in that area.

When actors are identified, the next step is to carry out research to understand the situation in the present. The study helps to know an initial state of data and characteristics of the target group, which is a reference point against which changes will be measured.

The most critical aspect in the success of the trials is to randomly select those who will enter into the treatment group and those who will go to the control group. Randomization, this being a main component of RCTs methodology, ensures that any differences which may be observed during analysis are attributed to the intervention and not pre-existing differences.

The group is allocated, then the intervention is implemented. At this point, the planned experimental intervention is done on the treatment group while the control group upholds the standard intervention or no intervention. After the intervention, a post-intervention survey will be conducted. This study evaluates the immediate effects of the intervention on participants by obtaining data on key variables identified as outcome measures at the beginning of the study.

Assessment of adoption rates forms an integral part of post-intervention analyses. It involves the evaluation of how many groups in the treatment group actually start using the new practices and techniques introduced by the intervention.

Now, the last stage involves a detailed analysis of the possible effects of the intervention on the control group. This analysis, along with comparing the pre- and post-intervention data, also evaluates the differences between the treatment and control groups to assess the effectiveness of the intervention.

Getting feedback and evaluations more often plays an important part throughout the process. These feedbacks involve regular monitoring and evaluation. Ultimately, it helps refine the trial methodology by implementing the adjustments and modifications to the intervention and trial methodology. The flexible nature of RCTs makes these trials trustworthy and highly adaptive practices that make sure the results are reliable and impactful. Therefore, it can have a great positive impact and help make informed decisions for future agricultural extension efforts.

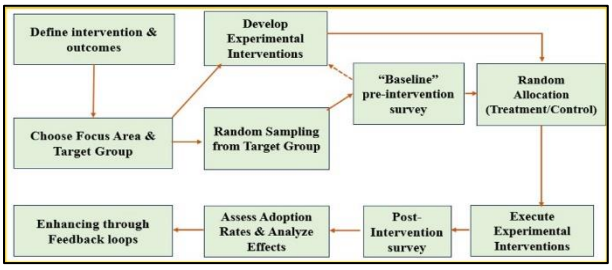


Figure (1): Sequential Steps in an RCT's Design and Implementation.

Source: Adapted from "Designing and Conducting Randomized Controlled Trials (RCTs) for Impact Evaluations of Agricultural Development" by [15] (p. 6)

Critical Arguments Against RCTs

Along with these positive sides, RCTs have many cons that need attention and further improvements to overcome these challenges and formulate more adaptable strategies. Sometimes, randomization doesn't completely eliminate biases, which can affect how we estimate the effects of treatment. Therefore, the results are not considered to represent the

broader populations, especially when significant heterogeneity exists within the sample. The effectiveness of the results normally depends on the context, with variations observed across different settings and populations. Critics also highlight the point that RCTs might just confirm predictable results, like the effectiveness of fertilizers on agricultural yields [1].

Methodological challenges such as attrition, partial compliance, contamination, diffusion, and spillover effects further threaten the validity of RCTs findings [17]. Ethical considerations, particularly in studies involving disadvantaged populations in which not all subjects receive the intervention, pose significant concerns [31, 32].

[31] contend that RCTs should not be uniquely prioritized for causal inference. They suggest that the choice of methodology should be guided by the particular research question, the existing body of knowledge, and the objectives of the study.

Designing an RCTs Study

Formulating the Research Question

Formulating a research question with population, intervention, comparator, outcome, and time (PICOT) elements is vital for clarity and specificity. For example, the question "Does the new fertilizer enhance crop yield?" can be refined using these elements. It is important to define the population (e.g., the crop or specific farmer demographic being studied), detail the intervention (the new fertilizer's type, application method, and frequency), identify the comparator (standard fertilizer or no fertilizer), specify the outcome (quantitative measure of crop yield increase), and determine the time (duration of the growing season or observation period post-fertilizer application). This approach ensures that a research question is well-structured and facilitates an effective investigation [19] (p. 382).

Study Objective

A study's objective should be quantifiable, meaning it can be evaluated using statistical methods. Many studies have multiple objectives. A common method for addressing this situation involves calculating the sample size needed for each objective separately. Then, the largest of these figures is chosen as the study's minimum required sample size [33]. In [33] (p. 18), the researcher proposes calculating the minimum sample size based solely on the primary objective of the research. This approach is deemed suitable when the primary objective carries greater importance than other objectives.

Clearly defining the target population, randomization process, allocation, inclusion and exclusion criteria, treatment and control delivery, intervention blinding, outcomes, required sample size, consent process, outcome assessment, ethical requirements, and data management are the next steps in an RCTs after a research question has been developed. These aspects of a study must be detailed in a well-structured protocol before the trial starts. In this article, we illuminate the essential components of designing an RCTs, as discussed in the following sections.

Select Appropriate Statistical Analysis or Methods

Detailed guidelines for planning RCTs are available in literature, and adhering to them is essential. The first step is to evaluate if an RCT is the most suitable method for the research question. The next stage is to carry out a thorough review of the literature to make sure that the topic hasn't already been addressed by previous, powerful RCTs. Then, it must be confirmed that the study's objectives can be met through statistical analysis, and the appropriate statistical methods to achieve these objectives are then chosen [33] (p. 23).

Pre-setting and Defining Power and Error Rates

Power: The statistical power of a hypothesis test refers to the probability of correctly detecting a true effect when it exists. It is influenced by the effect size, with larger effects being easier to detect. While increasing the significance level (alpha) can enhance power, alpha is typically kept at a conventional level (e.g., 0.05) to control for Type I errors (false positives). As a result, efforts to reduce Type II errors (false negatives) should focus on other strategies, such as increasing sample size, refining the experimental design, and ensuring consistency in implementation. These methods help boost power without altering the fixed Type I error rate. Notably, increasing sample size or raising the significance level can both substantially reduce the likelihood of a Type II error [21] (p. 46).

Type I error: When the null hypothesis is incorrectly rejected, falsely implying statistical significance because of chance or other circumstances, this is known as a Type I error.

The risk, denoted as alpha (α), is typically set at 5% (0.05), meaning there is a 5% chance the results could occur if the null hypothesis is true. Results with a p-value lower than α suggest significance, while a higher p-value indicates non-significance [21] (p. 44).The shaded tail end in Figure 2 signifies alpha. Results within this part of the curve are deemed statistically significant, leading to the rejection of the null hypothesis.

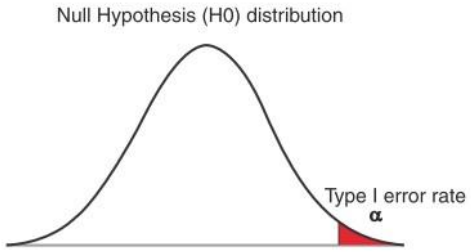


Figure (2): Null Hypothesis (H_0) Distribution
Source: Russo et al., 2022.

Type II error: A Type II error occurs when the null hypothesis is incorrectly accepted as true, meaning an actual effect is overlooked. This frequently occurs when a study lacks the statistical power necessary to identify a true effect. Acceptable power levels are usually 80% or higher, and as statistical power increases, the risk of a Type II error decreases [21] (p. 46)

The Type II error rate (β) is represented by the left-side blue area in the statistical distributions depicted in Figure 3. The area under the curve, minus this blue section, indicates the statistical power (1—β). Enhancing a test's statistical power reduces the likelihood of a Type II error [21] (p. 46).

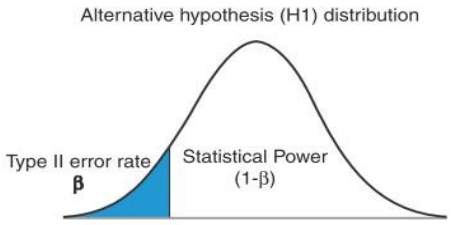


Figure (3): Alternate Hypothesis (H_1) Distribution.
Source: Russo et al., 2022.

The significance level, or Type I error rate, affects statistical power, that is inversely correlated with the Type II error rate. As a result, Type I and Type II errors are connected. Consequently, a significant trade-off exists: Type I mistake risk is decreased but Type II error risk is increased with a lower significance level. The risk of Type II errors is reduced when test power is increased.

Figure 4 illustrates the trade-off between Type I and also Type II errors. Hypothesis distributions overlap, creating an error area divided by type.

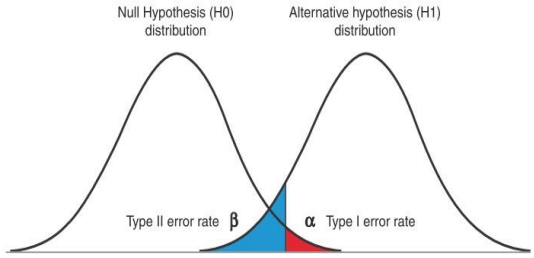


Figure (4): Error Trade off
Source: Russo et al., 2022.

The graph illustrates the trade-off between Type I and also Type II errors [21] (p. 47) Hypothesis distributions overlap, creating an error area divided into Type I and Type II errors.

Adjusting the Type I error rate (alpha) indirectly affects the Type II error rate (beta); decreasing alpha increases beta and vice versa, as shown in Figure 4.

Neither Type I nor Type II errors are inherently worse than the other type, and both carry significant consequences. Type I errors can lead to the adoption of ineffective policies or treatments and wasted resources. Type II errors risk overlooking beneficial new treatments or innovations [21] (p. 47).

The primary variable that remains to be determined in the sample size calculation is the effect size of the study [33] (p. 16). Understanding the type of study outcomes is essential to calculating effect size. In the ensuing sections, we will first explore different outcomes and then delve into definitions and methods for calculating effect size.

Definition and Identification of Outcomes

Based on [34], "outcomes," also known as events or endpoints, are key variables observed to assess the impact of certain interventions or exposures on agricultural practices. Typical outcomes in agricultural extension are increases in crop performance, better farming practices and economic growth. The "primary out-come" is the main variable that discloses on the research topic. It usually should be the feature that is the most significant for agricultural development of the welfare of the farmers. The secondary outcomes are the additional variables that are watched just to help understand the variations of the primary outcome [34]

To achieve an accurate review, the result has to be specified in quantitative terms. To illustrate, rather than a general objective, an exact yield increase, a rise in crop yield by 15% or a particular output magnitude in tons per hectare, should be specified. At this point, specifying targets in outline, for example, increasing wheat production to 2 from 2. 3 tons per hectare is a major factor that determines the accuracy of agricultural interventions.

In agricultural extension research which is about technology adoption or other interventions, the type of outcome (whether it is a continuous variable like yield improvement or a proportional variable like the percentage of farmers who adopt a technology) significantly affects how one calculates sample size to ensure the study has a sufficient power to find true effects.

This is because the variability and distribution of different types of outcomes can influence the statistical power of a study, which dictates the sample size needed to reliably observe an effect if one exists. For example, continuous outcomes with high variability might require a larger sample size for a specific change to be detected. Meanwhile, for proportional outcomes, especially if the expected proportion is very low or very high, there may be differences in the sample size needed to obtain the same degree of statistical precision.

Effect Size Planning

Effect Size Definition: According to [35] (p. 4), an effect size is a quantitative measure that reflects the magnitude of a phenomenon and offers a precise metric of the effect within a research context.

Cohen's Categories for Effect Size: Cohen defined a small effect size as 0.20, which is typically considered modest in research contexts. A medium effect size, noted as 0.50, reflects a more discernible and significant impact. Finally, 0.80 is considered a large effect size, which signifies a considerable and distinct difference in outcomes, according to [36] and [37] (p. 3).

Importance of Effect Size for Sample Size Calculation: [38] (p. 56) highlights that the anticipated effect size plays a vital role in determining a study's sample size. Whereas a smaller effect size necessitates a bigger sample to have the same statistical power, a greater effect size permits an effect to be observed with a smaller sample.

Effect Size Calculation by Means: Cohen's d is typically used to calculate effect size based on mean differences:

Formula: $d = \frac{(M1 - M2)}{SD_{pooled}}$ (1)

M_1 and M_2: the mean values for two groups.

SD pooled: the pooled standard deviation for the two groups. Formula is $\sqrt{[(S_1^2 + S_2^2)/2]}$. (2)

Cohen's d is a measure addressing the challenge of interpreting differences in means by considering variation within samples. It is one of the most recognized and widely utilized standardized effect sizes [39] and [40] (p. 722).

Effect Size Calculation Based on Proportions: Cohen's h is also used for proportions:

Formula: $h = 2\,[\arcsin(\sqrt{p1}) - \arcsin(\sqrt{p2})]$ (3)

p_1 and p_2: the proportions in each group.

This formula calculates the standardized difference between two proportions. It is useful in studies in which the outcomes are expressed as percentages or probabilities. The possible values for [*h*] range from 0 to π [40] (p. 719).

Power Calculation Procedures

Power calculations are crucial for determining adequate sample sizes and, in turn, preventing Type II errors, which inaccurately conclude a lack of program impact. This process involves assessing the effects of non-clustered and clustered program designs, identifying key outcome indicators linked to the program's objectives, establishing minimum impact levels justifiable by the program's costs and potential benefits, and calculating the baseline mean and variance of these indicators. Additionally, appropriate levels of statistical power (typically 0.8 or 0.9) and significance (commonly 5%) are essential to ensure robustness in detecting true effects. All these elements are fundamental to tailoring power calculations to specific evaluation needs and are supported by standard statistical software [17].

In RCTs, selecting the right sample size is crucial for a study's validity and reliability. The sample size must be adequately large for unbiased and precise results to be achieved. The determination of this optimal size involves specific calculations guided by the minimum detectable effect size formula. As outlined by [38], this formula is instrumental in addressing the query, "If my budget limits me to sampling only x households, what is the smallest effect size that can be reliably distinguished from no effect at all?" as cited in [2] (p. 255).

The minimum detectable effect size formula is expressed as follows:

$$\mathbf{MDE} = \left(t_{(1-k)} + t_{\alpha}\right)\sqrt{\frac{1}{P(1-P)}} * \sqrt{\left[\frac{\sigma^2}{N}\right]} \qquad (4)$$

In this formula:

$t_{(1-k)}$: statistical power

t_{α}: significance level

P: proportion of the treatment group within the sample

δ^2: variance

N: sample size

The MDE is conceptualized as either the anticipated effect size produced by the intervention or the threshold impact level below which the program is deemed ineffective. As detailed by [39], this formulation is integral to the design and interpretation of RCTs, as cited in [2] (p. 255).

Specialized software tools such as STATA are utilized to compute the sample size. In STATA, the specific command for this purpose is `. Power two proportions (pre) (post), m1() m2() rho () `, which effectively facilitates the calculation. In Appendix C, specifically in Example C1, a detailed calculation sheet is provided for practical reference.

Power Calculation for Clustered Trials:

The previous section covered power calculations for non-clustered programs. For programs with benefits assigned at the cluster level, the same five principles apply, with an added question:

Intra-cluster Correlation (ρ)

How variable is the outcome indicator within clusters, or what is the intra-cluster correlation (ρ)? High intra-cluster correlation requires more clusters for sufficient power. When randomization occurs at a cluster level, the intra-cluster correlation (ρ) and the average cluster size (m) must be incorporated into the denominator of the MDE in the formula. This adjustment is crucial for accurate calculations in cluster-level randomization scenarios [1].

$$\frac{\mathbf{MDE}}{\sqrt{1+\rho(m-1)}} = \left(t_{(1-k)} + t_{\alpha}\right)\sqrt{\frac{1}{P(1-P)}} * \sqrt{\left[\frac{\sigma^2}{N}\right]} \qquad (5)$$

Adding clusters vs. Adding individuals: Adding new clusters yields more power than adding individuals within existing clusters, though it can increase operational costs and complexity in data collection and program implementation.

Complexities in Power Calculations

Complexities in power calculations arise in practical, real-world scenarios beyond the basics of randomized assignment and complete compliance. For instance, quasi-experimental methods such as regression discontinuity, matching, and difference-in-differences typically require large sample sizes. In cases like regression discontinuity, a substantial sample around the eligibility threshold is crucial, and having multiple rounds of data can significantly enhance evaluation power. Moreover, when assessing varied program modalities or innovative approaches, the necessary sample size may differ based on specific policy questions. Sometimes, smaller effects are acceptable within treatment groups as opposed to comparisons between treatment and control groups. In multi-arm evaluations or when subgroup differences based on, for example, gender, age, or income are evaluated, sample size requirements can double, particularly when multiple subgroups are involved. Adjusted power calculations are essential in these scenarios to ensure sufficient power, and methods like stratified or block randomization can enhance power without increasing the sample size [17].

Resource Allocation for Handling Non-responses

Adjusting for Anticipated Drop-Out Rates: The required sample size should be increased to adjust for an anticipated dropout rate of r% in a study. This is done by multiplying the initial (unadjusted) sample size by the factor 100/(100 - r), as outlined by [41] (p. 330)[(1)] This formula accounts for potential losses to follow-up, ensuring that a study retains sufficient power [37] (p. 24). To prevent underestimating the sample size, researchers must consider potential non-response issues. Researchers should think about increasing the number of individuals above the minimal sample size by around 20% to 30% to make up for non-response problems [33] (p. 20).

Adjusting Allocation Ratios in Trials: In agricultural extension programs, these groups are likely to be equally supported, but sometimes certain situations call for uneven ratios. Taking a new farming technique that requires a high initial investment or specialized training for example. In this case, the intervention probably would be delivered to a smaller number of group than to the control group. Such an approach balances cost and resources and provides an opportunity to assess the impacts of this tech on crop yield and farming efficacy. Moreover, utilization of more controls than cases can elevate a case-control study's power. Sample size adjustment for different allocation ratios is derived using known formulas [42, 43] and also with the use of certain available software tools. Typically, the allocation ratio is

(1) If n is the sample size required per the formula and d is the dropout rate, then the adjusted sample size N1 is calculated as N1 = n/(1-d), as explained by [41].

set at 1:1, therefore, each group gets the same number of participants.

Software Tools for Sample Size Calculation in RCTs

Accurate determination of sample size is paramount in the design of Randomized Controlled Trials (RCTs), as it directly affects the study's statistical power, validity, and reliability. For researchers—especially those new to the field—the selection of appropriate software for sample size calculation can be daunting due to the myriads of available options and the statistical complexities involved. This section highlights widely utilized software tools for sample size calculation in agricultural RCTs.

When choosing software, researchers should consider the complexity of their study design, their level of statistical expertise, cost considerations, compatibility with existing analytical tools, and the availability of support and resources. Consulting experienced colleagues or statisticians can provide valuable guidance in the selection process. Refer to Appendix D for a comprehensive summary of these software options, including descriptions, key features, references, and download links. By carefully selecting the appropriate tool, researchers can ensure precise sample size calculations, thereby enhancing the rigor and success of their RCTs in agricultural research.

Writing a Sample Size Statement

The following instructions will help you prepare a sample size statement. It gives reasons why a particular number of subjects or units have been selected and demonstrates how the choice complies with the study's aims, power needs, and expected effect sizes. Thus, the rigorous planning for adequate sample size, which is crucial for the achievement of the statistically significant results, is highlighted in this thorough explanation, which underlines the importance of the scientific findings.

“The study hypothesized that a new dissemination approach significantly outperforms the conventional Training and Visit (T&V) method in encouraging the adoption of climate mitigation technologies over a one-year follow-up. To test this hypothesis, we determined the sample size using a two-sample proportions test within a clustered randomized design, employing Pearson's chi-squared test. Our estimation of the sample size is based on anticipated outcome proportions of 1% in the control group and 17.1% and 20% in the treatment groups, informed by prior research conducted by [44, 45], respectively. To dismiss the null hypothesis with 95% confidence and attain 80% power, each group requires a minimum of 10 and 8 participants, respectively. Accounting for an anticipated 20% dropout rate elevates the necessary sample to 12 and 10 participants per group.” This sample size statement has been refined and adapted, following the guidance outlined by [33] on pages 20 and 21.

Ethical Consideration

Interventions with known benefits should not be withheld from groups merely to conduct an evaluation [17] (p. 233). The careful observance of ethical guidelines is required for all RCTs involving trials on humans, animals, or human biological components [46]. Assessment of risks and benefits for individuals and society, ethical approval, and informed consent are crucial. Before an RCT is planned and executed, a careful evaluation is needed to determine the ethical appropriateness of using randomization to assign participants to intervention groups [19] (385). According to [47] (p. 260), informed consent and confidentiality are key ethical considerations in human research. Informed consent ensures participants' awareness of an agreement with the study, while confidentiality safeguards their personal data.

Blinding

Eliminating bias is crucial. Implementing single-blind or double-blind procedures, according to which participants (and perhaps the investigators) are unaware of group assignments, helps mitigate unconscious bias. However, the nature of some RCTs may preclude blinding [19] (p. 383).

Randomization Process

In RCTs, data is collected after the optimal sample size has been determined. The two-step method involves a baseline survey and randomization, which enhances precision but is costly and time-consuming. The one-step approach is simpler and quicker, but it entails direct randomization without baseline comparison, potentially sacrificing precision [1].

Level of Randomization: Randomization in research can occur at individual or cluster (like villages) levels and may be stratified. The choice of the type of randomization depends on budget, treatment spillover risks, and objectives, making the decision context-specific [18]. According to [19], cluster randomization is employed when individual participant randomization is impractical. In such scenarios, larger units like hospitals, clinics, or geographic areas are allocated as intervention or control groups. Field randomization may involve a coin toss, while office-based randomization uses software such as Stata or R [48].

Random Sampling Versus Random Allocation

The randomization process in an RCTs begins with the target population from which a random sample is drawn, ensuring external validity (the degree to which the results are applicable to a larger population). Those not selected for the evaluation sample are excluded from further analysis. Internal validity, which refers to the extent to which a study provides a causal relationship between an intervention and an observed outcome, is upheld by randomly dividing the selected evaluation sample into treatment and control groups. This random allocation is critical to reduce selection bias and confounding variables.

Random sampling and random allocation are distinct yet complementary strategies. Random sampling secures a study's relevance to the general population, enhancing the researchers’ ability to generalize results. Meanwhile, random allocation ensures unbiased comparisons within the study, preserving its internal validity. In essence, random sampling extends the study's findings to the broader population, while random allocation guarantees fairness in testing the intervention's efficacy [49, 50].

Method for Randomizing Treatment Allocation:

Randomized treatment assignment is vital for effective impact evaluations. It involves several critical steps, as described in the following sections.

Define Eligible Units

This process involves determining which entities (like an individual, a village, etc.) qualify for program participation based on specific criteria. Then, the size of the eligible population is evaluated and compared with the necessary number of observations for a thorough evaluation. If the eligible population

is small or exceeds the number needed, power calculations are used to ensure the accuracy of the evaluation.

Select a Representative Sample

In cases where the eligible population is extensive, a representative subset is chosen to effectively manage data collection costs and logistical constraints.

Randomized Assignment of Treatment and Comparison Groups.

For an equal 50-50 split, a coin should be flipped for each individual, with heads or tails predefined as indicating assignment to the treatment group. Dice rolls can be used to allocate precisely one-third of participants to the treatment group by assigning individuals to this group if a roll lands on 1 or 2. For larger samples, a random number generator or statistical software can be employed. It is crucial to establish and adhere to clear assignment rules and select participants with the highest random numbers for the treatment group. Transparency and integrity can be upheld by rigorously documenting the methodology and strictly following established rules, thus ensuring the randomization process is both transparent and verifiable [17] (pp. 77–78). Figure 5 illustrates the methodology for assigning participants to treatment and control groups using random values generated in an Excel spreadsheet.

Title: Implementing Randomization for Treatment Allocation Using a Spreadsheet

Randomization number	Between 0 and 1
Objective	Allocate 50% of evaluation sample to intervention group
Criteria	* If randomization number exceeds 0.5; include in the intervention group * If not, include in the control group

Identifier	Participant	Initial Random Value*	Final Random Value**	Group Assignment
1001	Ahmed	0.447326507	0.85989148	0
1002	Rosa	0.516043067	0.455289446	1
1003	Olivia	0.893539791	0.710017285	1
1004	Umar	0.541356359	0.451488746	1
1005	Vivian	0.560569791	0.629218765	1
1006	Wendy	0.621164544	0.472174521	1
1007	Xavier	0.57863653	0.267595454	1
1008	Yara	0.457440457	0.050661229	0
1009	Zoe	0.777993104	0.124379992	1
1010	Aisha	0.403234787	0.190888317	0
1011	Carlos	0.510736857	0.995289504	1
1012	Felix	0.433163123	0.926722175	0
1013	Diana	0.151734499	0.570848426	0

* **Action:** type the formula =RAND(). Reminder: The values in Column C are dynamic; they update with each new calculation.

** **Action**: Transfer values from Column C to Column D using "Paste Special> Values". Column D solidifies the random values

*** Formula: Enter =If(C(current row number)>0.5,1,0)

Figure (5): Spreadsheet Method for Random Allocation.
Source: adapted from [17], p. 78

Intervention Treatment

This crucial step involves an in-depth elaboration of the interventions or treatments to be applied during the study, as well as the detailing of the control settings for comparison. It encompasses the following components:

Treatment Description: This involves providing a comprehensive account of the interventions, including their theoretical basis, the rationale for their selection, and the methods of their implementation. This description should cover the duration, frequency, and intensity of each treatment to ensure replicability and transparency.

Control Settings: This step should clearly outline the nature of the control group settings. This may involve standard practice, no intervention, or a placebo treatment, depending on the study's design. In order to know how the intervention affected the outcome, the control setting must be defined.

Assessing Group Equivalence (The Balance Test)

In randomized studies, the balance test examines if the treatment and control groups are statistically similar in aspects like age, income, and land size. A lack of significant differences signifies effective randomization, leading to the formation of two comparable groups, as noted by [1] (p. 257). However, uniform differences in direction suggest a random assignment failure. [51] highlights the value of balance testing and questions the sufficiency of individual t-tests. The researcher also advocates for a joint orthogonality test using a regression model, [Treat = a + b1*X1 + b2*X2 + ... + b20*X20 + u], and tests the hypothesis [b1 = b2 = … = b20 = 0] with an F-test or chi-squared test in linear regression or probit models. Accepting the F-test's null hypothesis indicates successful balance. [52] argue that balancing observed variables implies a likely balance in unobserved variables under pure randomization. For additional details, refer to Box C1 in Appendix C.

Treatment Effect Estimation Techniques

The importance of participant selection rules cannot be overstated when selecting impact evaluation methods for social programs. Evaluation strategies should emphasize fairness and transparency and align closely with a program's operational policies. RCTs stand out in this regard, as they offer an equitable means of resource distribution and are recognized as the most effective method for measuring program impact [17] (pp. 63–64).

[17] The difference between the outcome with the program (Y| P = 1) and without it (Y| P = 0) is the fundamental formula for calculating the causal effect (Δ) of a program (P) on an outcome (Y). $\boldsymbol{Yi = \beta iTi + \sum_{j=1}^{J} \gamma iXij}$ is the regression equation used for estimation of average treatment effect? Ti is the treatment dummy variable (1 if treated, 0 otherwise), βi is the individual treatment effect, and xij represents the observed or unobserved covariates. [31] discuss the assumption that these covariates capture a sufficient set of outcome causes. Additionally, [53] notes that difference-in-differences or ANCOVA can be used for baseline data to estimate intent to treat (ITT) estimates.

Assessment Techniques in Non-Compliant Scenarios: Discerning imperfect compliance is critical when evaluating an RCT's impact. This circumstance often results in intention-to-treat estimates. Nevertheless, the local average treatment effect can be extracted using an instrumental variable method (IV)[(1)], notably a binary dummy for treatment allocation[(2)]. This method employs a two-stage least squares regression[(3)] to adjust for biases due to non-compliance, thereby uncovering the actual effect on recipients [17] (p. 162). The choice of regression model in the second stage, least squares regression analysis (2SLS), depends on the nature and distribution of the outcome variable, not on the 2SLS technique. The definitions of IV and 2SLS can be found on pages 34 and 35, respectively.

(1) An instrumental variable (IV) can be used to estimate the LATE. In this case, the IV is the original random assignment to the treatment group. It is correlated with the actual receipt of the training (due to random assignment) but is assumed to be independent of other factors that could affect crop yields.

(2) A binary-coded dummy variable is used, taking a value of 1 for units assigned to the treatment group and 0 for units assigned to the comparison group.

(3) Two-Stage Regression: A two-stage regression can be performed when using the IV. The first stage's purpose is to predict the actual receipt of the training based on the random assignment (the IV). In the second stage, the predicted values of training receipt are used to estimate the impact on crop yields. This method introduces bias by correcting imperfect compliance, and the true effect of the training on those who received it can be calculated.

The application of the 2SLS technique, combined with an IV strategy, effectively addresses the challenge of imperfect compliance. This approach refines the ITT methodology, providing a nuanced estimate of the treatment's impact on participants who adhere to their assigned treatment protocol. While ANCOVA serves to correct baseline disparities and improve the precision of ITT analysis, the integration of 2SLS with IV targets resolves issues related to participant non-compliance.

Evaluations typically compare outcomes between treatment and comparison groups to elucidate program effects. In ideal scenarios, they assess the average treatment effect under full compliance, by which everyone who is offered the program participates. However, real-world settings often exhibit partial compliance, leading to voluntary enrollment. Evaluations then measure the ITT effect, reflecting the impact of offering a program and the treatment-on-the-treated (TOT) effect, focusing on actual participants. ITT and TOT align under full compliance but differ otherwise, revealing insights into both the overall program offerings and specific impacts on participants [17] (pp. 90–91).

Interpreting RCTs Results

When interpreting RCTs in agricultural extension, statistical significance ($p < 0.05$) and practical relevance of outcomes should be assessed to ensure a sufficient sample size for detecting agriculturally meaningful differences. Small yet significant results might lack practical importance, while large, non-significant findings suggest an underpowered study. Moreover, applicability to real-world farming scenarios should be considered, and all evidence, including primary, secondary, and safety outcomes, for comprehensive insights should be evaluated. Recent issues with non-reproducibility in peer-reviewed studies emphasize the need for rigorous methodology and validation in agricultural research [19] (p. 386).

Guidelines for Reporting RCTs

The effective reporting of RCTs requires the CONSORT statement guidelines, accessible at www.consort-statement.org, to be followed rigorously. This includes prioritizing intention-to-treat analysis, recognized as the gold standard, over per-protocol analysis, which tends to overestimate the effects of interventions. It is essential to clearly define the trial type (superiority, noninferiority, or equivalence). Moreover, the detailed reporting of baseline characteristics of the participant groups is vital, as is the strict minimization of protocol violations. In the realm of primary outcomes, the current trend leans toward the comprehensive reporting of effect sizes, complete with 95% confidence intervals. Additionally, the provision of exact, two-sided p-values is crucial for presenting a more nuanced and precise understanding of a study's findings [19] (p. 385).

An Overview of Potential Challenges and Strategies for Mitigation

Table 1 below presents the significant challenges encountered by researchers in the design phase of RCT-based studies, along with various mitigation strategies.

Table (1): Methodological Challenges and Mitigation Measures in RCT-Based Studies.

Methodological Challenge	Mitigation Measures	Reference
Heterogeneous treatment effects: variations in the response to a program among different recipients within a group, indicating diverse outcomes.	- Consider stratified sampling by subpopulations. - Address low-representation subgroups. - Ensure a sufficient sample size to make meaningful impact estimations within specific groups.	[17] [p. 159)
Unintended behavioral effects include the Hawthorne effect (behavior change under observation), John Henry effect (greater effort by untreated units), anticipation (behavior change in untreated units expecting future treatment), and substitution bias (self-sought alternatives by non-chosen units).	- Use additional unaffected comparison groups as controls, providing a baseline to isolate and assess behavioral changes in the primary comparison group to ensure accurate program evaluation. - Explicitly test for unintended responses, quantifying behavioral effects (e.g., the Hawthorne effect, anticipation) to enhance evaluation accuracy. - Gather qualitative data to understand behavioral changes, ensuring internal validity in evaluation results.	[17] (pp. 160–161)
Imperfect compliance: occurs due to non-participation, administrative errors, unintended enrollment, non-enforcement of eligibility, and selective migration based on treatment status.	- Utilize intention-to-treat estimates and the instrumental variable approach for accurate local treatment effect analysis. - Adjust for complier proportions. - Select suitable instrumental variables based on program type. - Ensure balanced compliance in the treatment and comparison groups. - Be mindful of the method's limitations and specific applicability in different scenarios.	[17] (pp. 161–163) [11] (pp. 11–12)

Methodological Challenge	Mitigation Measures	Reference
	- Limit random assignment to a few staff, build an organizational understanding of its importance, and use community-level interventions.	
Spillover effects: unintended positive or negative impacts of an intervention on nonparticipants, including externalities, social interactions, context equilibrium, and general equilibrium effects [17] (p. 163)	Mitigate evaluation spillovers in RCTs and knowledge-based programs using community interventions, spatial separations, and cluster sampling. Apply inverse probability weights for precise estimates, expand scope, monitor impacts continually, select comparison groups judiciously, anticipate spillover mechanisms, utilize quasi-experimental designs as a backup, and conduct extensive data collection. Using a placebo control design counteracts partial compliance and contamination in experiments by comparing placebo and treatment groups to accurately isolate the intervention's effect.	[11] (pp. 11–12) [15] (p. 7) [1] (p. 261)
Attrition bias: It occurs in evaluations when participants drop out over time and affects sample representation and balance between treatment and comparison groups, compromising both the external and internal validity of the study.	To assess attrition in follow-up surveys, first, baseline characteristics of remaining and dropped-out units should be compared for similarity. Then, evaluate whether attrition rates are similar in treatment and comparison groups. If significant differences arise, employ statistical corrections like inverse probability weighting to align follow-up data with the initial baseline. [33] suggests over-recruiting by 20–30% to counter potential attrition or non-responses. [54] effectively employed a probit model and treatment effect bounds estimation.	[17] (pp. 169–170) [33] (p. 20) [54] (p. 14) [1] (p. 261)
Timing and persistence of effects: It refers to how quickly and how long the impacts of a program last. This includes the time needed for administrative setup, behavioral changes, and institutional adjustments. The persistence aspect considers whether these changes continue after the program ends.	When designing evaluations, it is crucial to realistically assess how long it will take for a program to reach its full effectiveness. This often requires multiple follow-up surveys to be conducted over time, including after the program has concluded, to accurately gauge the program's long-term impact and the durability of its effects.	[17] (p. 171)
Unreported effect sizes and novel inquiries in emerging fields, coupled with Hayes and Bennett's (1999) observation of the tendency to overestimate effect sizes, pose significant challenges in determining sample sizes for RCTs.	[33] (p. 19) recommends estimating effect size from recent or pilot studies or secondary data, noting the latter's limitations for new interventions. [55] (p. 6) suggests overcoming the shortage of ICC data by reviewing the literature, consulting experts, promoting future ICC reporting, and analyzing secondary data, as outlined in Appendix Table 3. They also propose using sensitivity analyses with various ICC values in new studies. [33] (p. 21) advises using conservative effect size estimates from the literature, incorporating sensitivity analyses in power calculations, and increasing sample sizes to account for potential dropouts and noncompliance.	[33] (pp. 19–21) [55] (p. 6)
Choosing a limited number of focused tests instead of a broad array.	Challenging experimental designs limit the number of testable options; practitioners often aim to test many, which is impractical. Treatments require uniform implementation for comparability, necessitating precise manuals for extension services, agencies, and NGOs to ensure consistent messaging to farmers and households.	[15] (p. 7)
Ethical challenges: Ethical issues in RCT-based evaluations primarily arise from the need to justify why certain participants are denied access to services. All RCTs involving experiments on animals, humans, or human biological materials must adhere strictly to ethical principles [17, 46].	Randomized allocation is fair but indifferent to participants' needs. When resources are limited, it can be more ethical than first-come-first-serve methods. Gradual rollouts and time-delayed implementations lessen ethical concerns, and providing extra value to the control group addresses service denial issues. Monitoring effects for early study termination minimizes delays but increases the risk of false positives.	[11] (p. 5)
Addressing the generalizability issue: Generalizability is a common challenge in impact evaluations focusing on specific interventions in particular contexts.	Randomized evaluations impartially estimate intervention effectiveness. Conducting them in varied or new environments broadens their relevance. Understanding the intervention theory	[16]

Methodological Challenge	Mitigation Measures	Reference
	and local specifics enhances replication and informs policy decisions.	
Insufficient technical expertise: Development organizations often lack impact evaluation expertise, leading to misunderstandings about an RCT's design, random assignment, and program impact, resulting in evaluation planning issues.	Effective capacity gap management requires early, clear communication. Evaluators should educate staff on RCTs and evaluation principles, guide self-reflection on program goals, and ensure stakeholder involvement in design decisions, thereby avoiding expedited processes that bypass crucial insights.	[11] (pp. 6–7)
Equivalence in groups and statistical power: Community interventions often face small sample sizes, increasing the risk of chance correlations due to contextual variability and making it hard to discern true treatment impact from other influences.	Collecting baseline data and using optimal matching enhance statistical power and reduce bias. Focusing on outcome rate changes rather than absolute levels offers a clearer intervention impact, while randomization inference tests assess treatment effects within specific samples, minimizing sampling variability.	[11] (pp. 10–11)
Program fidelity: In a multi-site agricultural program, varying regional implementation intensities due to local factors can impact the observed treatment effects.	Fidelity and implementation measures address treatment variation [11] (p. 12–13). By addressing treatment variation, these measures aim to determine how differences in an intervention's implementation may affect the outcomes.	[11] (p. 12–13)
Understanding program mechanisms (black box issue): RCTs can show whether a program works and its impact but might not reveal how or why it works.	Program evaluation necessitates an understanding of the theory of action (i.e., how inputs yield outcomes). Documenting, observing, and measuring implementation and participant reactions clarify RCTs effects. Moreover, qualitative data offers insights into participant views and impacts.	[11] (pp. 12–13)
Eagerness to innovate: Enthusiasm for new ideas can complicate results interpretation, as mid-evaluation changes and multiple partners can obscure the sources of impact.	It is crucial to establish consistent and clear communication channels, rigorously look for any deviations from the planned protocol, and meticulously document all alterations to ensure transparency and clarity in the study's methodology and results.	[11] (pp. 13–14)
Managing external intervention interference in RCTs: RCTs in multi-organization settings face challenges from interventions influencing control areas and potentially confounding results.	Conduct environmental scanning of ongoing and planned interventions to inform RCTs design, establish continuous monitoring and documentation systems, and spread-out interventions across different geographic areas to reduce external impact on study outcomes.	[11] (p. 14)
The risk of overestimating intervention effectiveness due to unadjusted standard errors in cluster-sampled designs	Neglecting to adjust standard errors in cluster-sampled designs can inflate statistical power estimates. This may lead to erroneous conclusions about an intervention's effectiveness and the false attribution of success when it may not be effective.	[15] (p. 7)
Identifying subgroup impacts beyond average effects.	RCTs provide unbiased average effects of interventions. However, when policymakers need effectiveness data for specific subgroups, especially in equity-focused projects [15] (p. 7), it is crucial to have adequate subgroup samples and use suitable statistical techniques.	[15] (p. 7)
Matching techniques for hub variability	Assess and group hubs based on various factors and conditions to minimize variability within treatment groups.	[11] (p. 5)

Source: This summary is derived from a range of references listed in the final column of the table.

Summary of Relevant RCTs Studies in Agricultural Settings

The agricultural extension literature features various studies utilizing RCTs across diverse technologies and interventions. Table 2 presents 21 selected cases of RCTs from esteemed journals, illustrating the broad applications of this methodology. While these cases provide valuable insights, they represent only a portion of the broader application of RCTs in the field of agricultural extension research. The RCT-based studies show that they play a crucial role in evaluating impacts and informing policy development.

The agricultural extension literature includes various studies that employ RCTs across a wide range of technologies and interventions. Table 2 presents 21 selected cases from esteemed journals, demonstrating the diverse applications of this methodology. These cases offer valuable insights but represent only a fraction of the broader use of RCTs in agricultural extension research. RCT-based studies play a crucial role in evaluating impacts and informing policy development.

Table (2): Key Studies on the Use of RCTs in Agricultural Development

Category	Subject of the Study	References
(1). Technological Innovations	Impact of the Avaaj Otalo (AO) mobile phone-based agricultural assistance service and in-person extension sessions on smallholder cotton farmers in Gujarat, India, focusing on information sources, agricultural knowledge, input adoption, productivity, and profitability.	[56]
	Impact of ICT-mediated agricultural advice, including video instructions, IVR services, and SMS reminders, on maize cultivation practices and yields among farm households in eastern Uganda.	[57]
	Impact of the MahindiMaster app on farmers' beliefs about agricultural input returns and their resource allocation in Western Kenya, with a focus on lime usage and soil pH levels.	[58]
(2). Farmer Education & Knowledge Transfer	Impact of different types of extension training on the adoption of a biofortified bean variety (KK15) among smallholder farmers in Western Kenya.	[13]
	Effectiveness of different knowledge-transfer interventions on improving equid health knowledge among rural Ethiopian donkey owners, focusing on wound management.	[20]
	Impact of Farmer Field Schools (FFS) or similar agricultural extension programs on rice production practices, including fertilizer use, crop yields, income, and knowledge acquisition, among farmers in China. The study also examines how impacts vary by gender and age.	[26]
	Impact of enhanced training for development agents (DAs) and the use of demonstration trials/field days on the adoption of a new wheat variety, yields, and chemical input use among smallholder wheat farmers in Ethiopia.	[54]
	Impact of video documentaries (VDD) and radio listening clubs (RLCs) on the awareness, uptake of Brady-rhizobium inoculants, use of improved legume seeds, and legume yields among smallholder farm households in northern Ghana.	[59]
	Impact of a 12-week home garden-based intervention on vegetable consumption, self-efficacy, preferences, knowledge, and anthropometric measurements among healthy Malaysian children aged nine to 12 living in low-cost housing flats in Kuala Lumpur.	[60]
	Impact of farmer field school (FFS) training on urban and peri-urban vegetable producers' knowledge, practices, and awareness of integrated pest management (IPM) and pesticide use in Cotonou, Benin	[61]
(3). Digital vs Traditional Extension Methods	Impact of the Digital Green system versus traditional agricultural extension services on the adoption of agricultural practices among farmers in 16 villages in India.	[62]
	Effectiveness of different knowledge-dissemination techniques—namely diagrammatic handouts, village meetings, videos, and their combinations—in enhancing mastitis knowledge among Tanzanian smallholder dairy farmers.	[45]
	Impact of digital extension services, specifically site-specific nutrient management recommendations, on maize farmers' fertilizer use, yields, and income in northern Nigeria.	[63]
(**4). Crop and Livestock Management**	Impacts of diverse training on barley adoption, agricultural practices, livelihoods, and gender empowerment among small ruminant farmers in Tunisia.	[44]
	Impact of aquaculture training on small-scale pond farmers in Ghana's six principal tilapia-producing regions, focusing on productivity, income, and adoption of improved practices.	[64]
	Impact of "farmer-to-farmer" extension activities and additional video interventions on the adoption and knowledge of Integrated Soil Fertility Management (ISFM) practices among smallholders in Ethiopia.	[65]

(5). Economic and Social Interventions	Impact of conventional subsidies versus agglomeration payments on the adoption of conservation agriculture (CA) practices among farmers in the Shire River Basin in southern Malawi.	[66]
	Impact of agricultural, nutrition, and marketing training sessions on the adoption of the black bean variety KK15 and communication networks among smallholder farmers in rural Kenya.	[67]
	Impact of various interventions—including conditional cash transfers, vocational training, business grants, and weather index insurance—on the ability of agricultural households in developing countries to cope with and adapt to agricultural risks and shocks, focusing on income diversification, risk management strategies, and investment behaviors.	[68]
(5). Economic and Social Interventions	Impact of providing soil-specific fertilizer information and input vouchers on fertilizer use, maize yields, and agricultural investment among smallholder farmers in Morogoro Rural, Tanzania.	[69]
	Impact of gender empowerment workshops on promoting shared responsibilities and enhancing women's roles in coffee-producing communities in Honduras. It measures outcomes in gender equity, leadership roles, decision-making authority, and satisfaction with leisure time.	[70]

Note: This table was put together by the author based on a thorough review of the literature. It highlights key studies using RCTs in agricultural development. For more detailed information on each study, please refer to the listed references.

Conclusion

Even though RCTs are a source of strong evidence in agricultural extension programs, their design and implementation demand careful consideration. Understanding the essential concepts, appropriate design, and inherent limitations of RCTs is vital, and agricultural professionals and researchers should be well-versed in these aspects before implementing RCTs. Adherence to rigorous protocols is necessary both in the application of RCTs and in the effective utilization of their results.

Owing to the growing use of cluster randomized trials in agricultural extension, it is crucial to focus on producing academic reports that are informative and aid in effectively interpreting such trials. This would ensure their practical and theoretical utility in the field.

This guide streamlines the process of designing and implementing RCT-based studies for agricultural extension professionals and researchers. It equips them with the tools needed to conduct methodologically robust RCTs studies, thereby significantly enhancing the quality and effectiveness of ongoing impact evaluation efforts in the field. Reading this guide will help understand the process's intricacy and RCTs study quality. This will improve how knowledge synthesis is included in the creation of policies and decision-making processes.

The outcomes of RCT-based studies in agricultural extension are pivotal in enhancing both the outreach and cost-efficiency of extension programs. If these findings are leveraged, extension services can be optimized to better serve the most underprivileged farmers, which would significantly improve their agricultural practices, livelihoods, and overall quality of life. These findings promise immediate benefits for farmers and long-term, sustainable growth and development in rural communities. Therefore, researchers need to increasingly utilize RCTs in the evaluation of agricultural extension services to further these gains and ensure the efficacy and impact of future interventions.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable

Author’s contribution

The first author, A.A., was responsible for the conception, design, and drafting of the manuscript. The co-authors, A.M. and H.B., reviewed the draft and provided valuable comments and suggestions. All authors have read and approved the final version of the manuscript.

Availability of data and materials

The article itself contains the facts that back up the conclusions of this investigation. On request, further information may be given.

Funding

No funding has been received for this work.

Conflicts of Interest:

Regarding the release of this work, the authors affirm that they have no conflicts of interest.

Acknowledgements

The authors are grateful to the two anonymous reviewers for their valuable feedback, insightful suggestions, and constructive comments, which have significantly improved the quality and clarity of this manuscript.

Key concepts

Bias and Validity Issues

Attrition: Loss of subjects leading to skewed impact estimates [16].

Selection Bias: Participation factors influencing outcomes, affecting estimated impact [17] (p. 59), differences due to pre-existing variations [16] (p. 55).

Confounding Factors: Variables influencing the outcome [16].

External Validity: Sample representing the broader population [17] (p. 73).

Internal Validity: Isolating the program's impact from confounding factors [17] (p. 71).

Evaluation Methods

Impact Evaluations: Determining the program's causal effect [17] (pp. 7–8).

Intention to Treat (ITT): Analyzing participants based on original group allocation [47] (p. 259).

Efficacy vs. Effectiveness Studies: Testing optimal conditions vs. real-world operation [17] (p. 11).

Analysis Techniques:

Cost-Benefit & Cost-Effectiveness Analysis: Comparing benefits vs. costs and expenses of multiple programs [17] (p. 18).

Counterfactual: Hypothetical outcome without the program [17] (p. 49).

Difference-in-Differences: Measuring outcome changes over time [17] (p. 130).

Instrumental Variable (IV): Estimating causal effects with confounding variables [71],[72].

Two-Stage Least Squares (2SLS): Extending OLS methodology for structural equation analysis [73].

Power Calculation: Detecting effect size within a narrow confidence interval [74] (p. 2).

Design and Sampling

Contamination: Control group influenced by the intervention [16] (p. 52).

Spillovers: Comparison group's results influenced by the treatment group [17] (p. 79).

Intracluster Correlation Coefficient (ICC): Outcome similarity within clusters [18, 54, 75-77]. (pp.68-70)

Matching: Creating an optimal comparison group using statistical methods [17] (p. 143).

Randomized Assignment: Ensuring equal chance of selection for the intervention [17] (pp. 64–68).

Sample: Accurately representing both study and target populations [78] (p. 143).

Conceptual Frameworks

Results Chain: Sequence from inputs to outputs enhancing outcomes [17, 79] (p. 34).

Theory of Change: Explaining the causal rationale and methods for achieving goals [17] (p. 32).

Unobservable Characteristics: Traits leading to selection bias [16].

References

1] Cariappa AG. An introduction to randomized evaluations and its application in agricultural economics. Agricultural Economics Research Review. 2021;34:253-65.

2] Birner R, Davis K, Pender J, Nkonya E, Anandajayasekeram P, Ekboir J, et al. From Best Practice to Best Fit: A Framework for Designing and Analyzing Pluralistic Agricultural Advisory Services Worldwide. The Journal of Agricultural Education and Extension. 2009;15(4):341-55.

3] Swanson BE, Ed.; Bentz, Robert P., Ed.; Sofranko, Andrew J., Ed editor. Improving Agricultural Extension: A Reference Manual. Rome: FAO; 1997.

4] Van den Ban A. Supporting farmers, decision making by agricultural extension. Journal of Extension Systems. 1998;14:55-67.

5] Davis K. Farmer Field Schools: A Boon or Bust for Extension in Africa? Journal of International Agricultural and Extension Education. 2006;13.

6] Dissanayake C, Jayathilake W, Wickramasuriya H, Dissanayake U, Wasala W. A Review on Factors Affecting Technology Adoption in Agricultural Sector. Journal of Agricultural Sciences–Sri Lanka. 2022;17(2):280-96.

7] Anderson JR, Feder G. Rural Extension Services. Washington, DC: Agriculture and Rural Development Department, World Bank; 2003. Word Bank Policy Research Working Paper 2976.

8] Feder G, Willett A, Zijp W. Agricultural Extension: Generic Challenges and the Ingredients for Solutions. In: Wolf SA, Zilberman D, editors. Knowledge Generation and Technical Change: Institutional Innovation in Agriculture. Boston, MA: Springer US; 2001. p. 313-53.

9] Ogundari K. A meta-analysis of the impact of agricultural extension services. China Agricultural Economic Review. 2022;14(2):221-41.

10] Aker JC. Dial “A” for agriculture: a review of information and communication technologies for agricultural extension in developing countries. Agr Econ. 2011;42(6):631-47.

11] Culbertson MJ, McCole D, McNamara PE. Practical challenges and strategies for randomised control trials in agricultural extension and other development programmes. Journal of Development Effectiveness. 2014;6(3):284-99.

12] Winters P, Maffioli A, Salazar L. Introduction to the Special Feature: Evaluating the Impact of Agricultural Projects in Developing Countries. Journal of Agricultural Economics. 2011;62(2):393-402.

13] Ogutu SO, Fongar A, Gödecke T, Jäckering L, Mwololo H, Njuguna M, et al. How to make farming and agricultural extension more nutrition-sensitive: evidence from a randomised controlled trial in Kenya. European Review of Agricultural Economics. 2018;47(1):95-118.

14] Scartascini CG. Let There Be (More) Light: The Case for Impact Evaluations [Internet]. Inter-American Development Bank (IDB); 2015 Jul 16 [cited 2024 Mar 16]. Available from: https://perma.cc/Y2CJ-QQ2U.

15] Dhehibi B, Werner J, Qaim M. 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: The International Center for Agricultural Research in the Dry Areas (ICARDA); 2018.

16] J-PAL. Introduction to Randomized Evaluations: Training at CIMMYT [Internet]. Abdul Latif Jameel Poverty Action Lab and Agricultural Technology adoption Intiative training at CIMMYT; 2012 [cited 2024 Mar 16]. Available from: https://perma.cc/89AA-LD9L.

17] Gertler PJ, Martinez S, Premand P, Rawlings LB, Vermeersch CMJ. Impact Evaluation in Practice. 2nd ed. Washington, DC: World Bank; 2016.

18] Duflo E, Glennerster R, Kremer M. Using randomization in development economics research: A toolkit. Handbook of development economics. 2007;4:3895-962.

19] Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. Acta Obstet Gynecol Scand. 2018;97(4):380-7.

20] Stringer AP, Bell CE, Christley RM, Gebreab F, Tefera G, Reed K, et al. A cluster-randomised controlled trial to compare the effectiveness of different knowledge-transfer interventions for rural working equid users in Ethiopia. Prev Vet Med. 2011;100(2):90-9.

21] Russo E, Montalti A, Santonastaso DP, Bolondi G. Randomized Trials and Case–Control Matching Techniques. In: Ceresoli M, Abu-Zidan FM, Staudenmayer KL, Catena F, Coccolini F, editors. Statistics and Research Methods for Acute Care and General Surgeons: Springer; 2022. p. 125-37.

22] Thomas DP. Sailors, scurvy and science. Journal of the Royal Society of Medicine. 1997;90(1):50-4.

23] Gibson M, Sautmann A. Introduction to randomized evaluations. Abdul Latif Jameel Poverty Action Lab. 2017.

24] Athey S, Imbens GW. The econometrics of randomized experiments. Handbook of economic field experiments. 1: Elsevier; 2017. p. 73-140.

25] Groenwold RHH, Van Deursen AMM, Hoes AW, Hak E. Poor Quality of Reporting Confounding Bias in Observational Intervention Studies: A Systematic Review. Ann Epidemiol. 2008;18(10):746-51.

26] Guo M, Jia X, Huang J, Kumar KB, Burger NE. Farmer field school and farmer knowledge acquisition in rice production: Experimental evaluation in China. Agriculture, Ecosystems & Environment. 2015;209:100-7.

27] Doss CR. Analyzing technology adoption using microstudies: limitations, challenges, and opportunities for improvement. Agr Econ. 2006;34(3):207-19.

28] Zabor EC, Kaizer AM, Hobbs BP. Randomized Controlled Trials. Chest. 2020;158(1s):S79-s87.

29] Gomez KA, Gomez AA. Statistical procedures for agricultural research. 2nd ed. New York: John Wiley & Sons 1984. 690 p.

30] Hebbar A, Suma A. R codes for experimental designs in agricultural research with special reference to CRD, RCBD, LSD, split-plot and strip-plot. 2024.

31] Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Social Science & Medicine. 2018;210:2-21.

32] Deaton A. Randomization in the tropics revisited: a theme and eleven variations. National Bureau of Economic Research; 2020.

33] Bujang MA. A Step-by-Step Process on Sample Size Determination for Medical Research. The Malaysian Journal of Medical Sciences : MJMS. 2021;28:15 - 27.

34] Ferreira JC, Patino CM. Types of outcomes in clinical research. J Bras Pneumol. 2017;43(1):5.

35] Kelley K, Preacher KJ. On effect size. Psychological Methods. 2012;17(2):16.

36] Cohen J. Statistical Power Analysis for Behavioral Sciences, Second edition. Lawrence Erlbaum Associates, Mahwah NJ.1987.

37] Malone H, Nicholl H, Coyne I. Fundamentals of Estimating sample size. Nurse Researcher. 2016;Volume 23:21-5.

38] Kadam P, Bhalerao S. Sample size calculation. Int J Ayurveda Res. 2010;1(1):55-7.

39] Glen S. Cohen's D: Definition, Examples, Formulas [Available from: https://perma.cc/M48D-4VWP

40] Dey R, Mulekar MS, editors. Effect Size as a Measure of Difference Between Two Populations. Encyclopedia of Social Network Analysis and Mining 2nd Ed; 2018.

41] Gupta KK, Attri JP, Singh A, Kaur H, Kaur G. Basic concepts for sample size calculation: Critical step for any clinical trials! Saudi J Anaesth. 2016;10(3):328-31.

42] Sakpal TV. Sample size estimation in clinical trial. Perspectives in clinical research. 2010;1(2):67.

43] Machin D, Campbell MJ, Tan SB, Tan SH. Sample size tables for clinical studies: John Wiley & Sons; 2009.

44] Dhehibi B, Dhraief MZ, Ruediger U, Frija A, Werner J, Straussberger L, et al. Impact of improved agricultural extension approaches on technology adoption: Evidence from a randomised controlled trial in rural Tunisia. Experimental Agriculture. 2022;58.

45] Bell C, French N, Karimuribo E, Ogden N, Bryant M, Swai E, et al. The effects of different knowledge-dissemination interventions on the mastitis knowledge of Tanzanian smallholder dairy farmers. Preventive veterinary medicine. 2005;72(3-4):237-51.

46] WMA. WMA Declaration of Helsinki: Ethical Principles for Medical Research Involkving Human Subjects [Internet]. 2022 [cited 2024 Mar 16]. Available from: https://perma.cc/XAT5-4PPY

47] Stanley K. Design of Randomized Controlled Trials. Circulation. 2007;115(9):1164-9.

48] WorldBank. Randomization in Stata. DIME Wiki, : World Bank; 2021 [Available from: https://perma.cc/WU69-ULTV.

49] Kim J, Shin W. How to do random allocation (randomization). Clin Orthop Surg. 2014;6(1):103-9.

50] Dunn PK. Random Sampling and Allocation. In: Statistical Concepts [Internet]. 2021 [cited 2024 Mar 16]. Available from: https://perma.cc/5K4U-FCZD.

51] McKenzie D. Development Impact [Internet]: World Bank Blogs. 2015 Feb 4 [cited 2024 Apr 24]. Available from: https://perma.cc/TKF2-TJ43.

52] Bruhn M, McKenzie D. In pursuit of balance: Randomization in practice in development field experiments. American economic journal: applied economics. 2009;1(4):200-32.

53] Özler B. World Bank Blogs [Internet]. 2015 [cited 2023 Jan 10]. Available from: https://blogs.worldbank.org/impactevaluations/why-difference-difference-estimation-still-so-popular-experimental-analysis.

54] Yitayew A, Abdulai A, Yigezu YA, Deneke TT, Kassie GT. Impact of agricultural extension services on the adoption of improved wheat variety in Ethiopia: A cluster randomized controlled trial. World Development. 2021;146:105605.

55] Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. International journal of epidemiology. 1999;28(2):319-26.

56] Cole SA, Fernando AN. 'Mobile'izing agricultural advice technology adoption diffusion and sustainability. The Economic Journal. 2021;131(633):192-219.

57] Van Campenhout B, Spielman DJ, Lecoutere E. Information and communication technologies to provide agricultural advice to smallholder farmers: Experimental evidence from Uganda. Amer J Agr Econ. 2021;103(1):317-37.

58] Tjernström E, Lybbert TJ, Hernández RF, Correa JS. Learning by (virtually) doing: Experimentation and belief updating in smallholder agriculture. Journal of Economic Behavior & Organization. 2021;189:28-50.

59] Dzanku FM, Osei RD, Nkegbe PK, Osei-Akoto I. Information delivery channels and agricultural technology uptake: experimental evidence from Ghana. European Review of Agricultural Economics. 2020;49(1):82-120.

60] Mok KT, Tung SEH, Kaur S, Chin YS, Martini MY, Ulaganathan V. A home gardening intervention to improve vegetable consumption among urban poor children: A study protocol for randomised controlled trial. Nutrition and Health. 2023;29(1):9-20.

61] Lund T, Sæthre M-G, Nyborg I, Coulibaly O, Rahman M. Farmer field school-IPM impacts on urban and peri-urban vegetable producers in Cotonou, Benin. International Journal of Tropical Insect Science. 2010;30(1):19-31.

62] Gandhi R, Veeraraghavan R, Toyama K, Ramprasad V, editors. Digital green: Participatory video for agricultural extension. 2007 International conference on information and communication technologies and development; 2007: IEEE.

63] Oyinbo O, Chamberlin J, Abdoulaye T, Maertens M. Digital extension, price risk, and farm performance: experimental evidence from Nigeria. Amer J Agr Econ. 2022;104(2):831-52.

64] Ragasa C, Amewu S, Agyakwah SK, Mensah ET-D, Asmah R. Impact of aquaculture training on farmers' income: Cluster randomized controlled trial evidence in Ghana. Agr Econ. 2022;53(S1):5-20.

65] Hörner D, Bouguen A, Frölich M, Wollni M. Knowledge and Adoption of Complex Agricultural Technologies: Evidence from an Extension Experiment. The World Bank Economic Review. 2021;36(1):68-90.

66] Ward PS, Mapemba L, Bell AR. Smart subsidies for sustainable soils: Evidence from a randomized controlled trial in southern Malawi. Journal of Environmental Economics and Management. 2021;110:102556.

67] Jackering L, Fongar A, Godecke T, Mbugua M, Njuguna M, Ogutu S, et al. Communication Networks and the Adoption of Technologies: Evidence from a Randomized Experiment. 2018// July.

68] Macours K. Volatility, agricultural risk, and household poverty: micro‐evidence from randomized control trials. Agr Econ. 2013;44(s1):79-84.

69] Harou AP, Madajewicz M, Magomba C, Michelson H, Tschirhart K, Amuri N, et al. Can information improve investment? Effects of site-specific soil recommendations on fertilizer demand. Unpublished mansucrip, McGill University. July 2018.

70] Rubio-Jovel K, editor Gender Empowerment in Agriculture Interventions: What Are We Still Missing? Evidence From a Randomized-Controlled Trial Among Coffee Producers in Honduras. Front Sustain Food Syst. 2021.

71] Bobbitt Z. Instrumental Variables: Definition & Examples [Internet]. 2020 [cited 2024 Apr 14]. Available from: https://perma.cc/GE4N-BW8R.

72] Columbia University MSoPH. Instrumental Variables [Internet]. [cited 2024 Apr 14]. Available from: https://perma.cc/M5QS-67JF.

73] Dissertation C. Two-Stage Least Squares (2SLS) Regression Analysis [Internet]. [cited 2024 Apr 14]. Available from: https://perma.cc/MH8G-M3HN.

74] Hemming K, Eldridge S, Forbes G, Weijer C, Taljaard M. How to design efficient cluster randomised trials. bmj. 2017;358.75] Collaboratory N. Cheat Sheet on the Intraclass Correlation Coefficient. Rthinking Clinical Trials 2020. Available from: https://perma.cc/GQ5K-24ES.

76] Campbell MJ, Donner A, Klar N. Developments in cluster randomized trials and Statistics in Medicine. Statistics in Medicine. 2007;26(1):2-19.

77] Young A. Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. The quarterly journal of economics. 2019;134(2):557-98.

78] Omair A. Sample size estimation and sampling techniques for selecting a representative sample. Journal of Health specialties. 2014;2(4):142.

79] Baslé M. Strengths and weaknesses of European Union policy evaluation methods: Ex-post evaluation of objective 2, 1994–99. Regional Studies. 2006;40(2):225-35.

80] FAO. Chapter 6: Completely Randomized Designs n.d. Available from: https://perma.cc/3EVM-HXCD.

81] Piquette NA, Savage RS, Abrami PC. A cluster randomized control field trial of the ABRACADABRA web-based reading technology: replication and extension of basic findings. Frontiers in Psychology. 2014;5:1413.

82] Campbell MK, Grimshaw JM, Elbourne DR. Intracluster correlation coefficients in cluster randomized trials: empirical insights into how should they be reported. BMC Medical Research Methodology. 2004;4(1):9.

83] Campbell M, Grimshaw J, Steen N. Sample Size Calculations for Cluster Randomised Trials. Journal of Health Services Research & Policy. 2000;5(1):12-6.

84] Campbell MK, Thomson S, Ramsay CR, MacLennan GS, Grimshaw JM. Sample size calculator for cluster randomized trials. Comput Biol Med. 2004;34(2):113-25.

85] Faul F, Erdfelder E, Lang A-G, Buchner A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods. 2007;39(2):175-91.

86] NCSS L. PASS: Power Analysis and Sample Size Software [Internet]. 2023. [cited 2024 Sep 17]. Available from: https://www.ncss.com/software/pass/.

87] StataCorp. Stata Statistical Software: Release 16. College Station, TX: Stata Corp LLC; 2019 [Internet]. n.d. [cited 2024 Sep16]. Available from: https://www.stata.com/.

88] IBM Corp. IBM SPSS Sample Power [Internet]. 2017. [cited 2024 Sep17]. Available from: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics 28010?mhsrc=ibmsearch_a&mhq=download%20spss.

89] Dean AG, Sullivan KM, Soe MM. OpenEpi: Open Source Epidemiologic Statistics for Public Health [Internet]. 2013. [cited 2024 Sep17]. Available from: http://wwwopenepicom/.

90] Statulator: An online statistical calculator [Internet]. n.d. [cited 2024 Sep17]. Available from: https://statulator.com.