FEED THE FUTURE
The U.S. Government's Global Hunger and Food Security Initiative



**Africa RISING**

**USAID's Sustainable Intensification Program in Africa**

# Impact Assessment of Africa RISING: Approaches and Challenges

IFPRI®

**March 6, 2013**

HarvestChoice
BETTER CHOICES, BETTER LIVES

## Table of Contents

**Impact Assessment of Africa RISING: Approaches and Challenges**

## 1. Introduction[1]

Africa RISING (AR) provides great opportunities to learn about what works and what does not work, along with its precise impact and causal mechanisms. Information collected as part of AR can support various types of evaluation, especially if evaluation designs are carefully considered at the outset of the program. USAID's evaluation policy also specifies an independent (and rigorous) evaluation, with the recognition that much valuable learning can also be achieved through evaluations carried out by implementers of the various projects.

This document discusses requirements for rigorous impact assessment of AR research activities and assesses whether such evaluation is feasible in the context of AR, given the scale and type of ongoing and planned research activities and available resources. While there are various quantitative and qualitative approaches to impact assessment and analysis of potential transmission mechanisms, this document explores requirements for a quantitative impact assessment through random assignment of units into treatment and control – Randomized Controlled Trials (RCTs)– and outlines other alternative/complementary approaches to RCTs.

Especially after the Accra meeting in January, the M&E team and AR implementing partners in the three mega-sites have been discussing about possible approaches to evaluate AR. Below are some of the issues/activities flagged/addressed through these discussions.

A. The scale of AR research activities and implications for evaluation approach
B. Clarity on (planned) activities
    - Project-specific work plan (project objectives, hypotheses, indicators, interventions, timing, number of expected beneficiaries, control over the research activities given multiple partnerships, etc.)

[1] This document is not supposed to be providing an exhaustive list of all the issues and challenges associated with M&E and impact assessment of Africa RISING program but rather to guide the discussion about impact assessment of the program at Lilongwe meeting. The reader is advised to refer to the AR M&E plan for detailed discussion of relevant issues.

C. Evaluation approach

- Is there a need for assessing the impact of activities in each focus country? What approach to pursue and to what end? What are the strengths and limitations of each approach?
  i. Experimental (simple RCT, phased-in RCT)
  ii. Non-experimental (requires stronger identifying assumptions)
  iii. Diffusion of innovations - intensive engagement with small number of farmers at the beginning to "generate evidence" for interventions later on
  iv. Farming systems analysis
  v. Qualitative evaluation (focus groups, participatory analysis)
  vi. A combination and, if so, which combination?

D. Survey

- Baseline survey design (influenced by the type of evaluation approach to be pursued)
- Survey timing (collection of baseline data for RO1 only, or also for impact assessment?)
- Survey design meetings (extremely helpful to better adapt draft questionnaire put together by IFPRI to the local context, suggestions made to include additional modules, concerns raised about the length of the draft instrument and the relevance of some modules and questions)
- Selection of study units (random versus non-random) and implications (selection bias, ethical concerns)
- Power calculation -large number of clusters with low number of farmers per cluster versus low numbers of cluster with large number of farmers per cluster
- Sampling frame needs

E. Other related issues

- Survey implementation - discussions with local entities to help IFPRI with data collection (Innovations for Poverty Action and ISSER in Ghana, Economic Development Initiatives in Tanzania, Agricultural Transformation Agency in Ethiopia)
- Tasks, roles, and responsibilities of local M&E coordinators (IFPRI made offers to 2 candidates for ESA and WA sites on 3/1/2013, awaiting their response)
- Office station for local M&E coordinators in Tamale and Arusha (lease just signed by IITA for a IITA office in Tamale, and office space in IITA at the AVRDC compound in Arusha)

While the AR team is expected to have more clarity and agreement on AR research activities in general and items (A) (B) in particular at the Malawi meeting, the M&E team strongly feels that the team should also reach a consensus about the evaluation approach to be pursued (C) so that it can proceed with the data collection process (D) as soon as possible in Northern Ghana (and eventually Southern Mali); right after the May-June harvest in Tanzania (and eventually in Malawi) ); and right after the September harvest in the Ethiopian Highlands.

The rest of the document is organized as follows. Section 2 discusses the requirement for an RCT approach, as it relates with survey design, sample selection process, timing, and sample size calculation. Section 3 outlines alternative/complementary approaches to RCT. Section 4 concludes the document with a note about IFPRI's position on the impact evaluation approach.

## 2. Requirements for impact assessment through Randomized Controlled Trials (RCT)

**Random selection** - An important aspect of an RCT approach is the ability to randomize units (villages, households, etc.) into treatment and control.[2] With proper evaluation design,

[2] For example, in the Cereal Systems Initiative for South Asia (CSISA) project that encompasses nine hubs in four South Asian countries (5 hubs in India, 1 hub in Nepal, 2 hubs in Bangladesh, and 1 hub in Pakistan), 18 villages from each hub (9 intervention and 9 control) and 18 households from each village (for a total final sample of 2,628 households) were randomly sampled for a baseline survey in 2010/11 to establish a priori conditions against which the social, economic, and livelihood impacts of CSISA would be evaluated.

randomized selection of action research sites and farmers coupled with selection of controls that are isolated from research sites but have characteristics that are as much similar to the research sites as possible would provide statistically robust estimates of program impact that are free from selection biases.[3] Due to ethical concerns associated with simple RCT, there has been a discussion about a phased-in RCT approach to assess program impact when research activities are rolled out to sites previously classified as controls, suggesting a pipeline evaluation approach[4]. The below table provides an example a phased-in RCT.

| Community | Project Year/ Treatment Status | | |
|---|---|---|---|
| | Time=t | Time=t+1 | Time=t+2 |
| Group 1 (T1) | treated | treated | treated |
| Group 2 (T2) | control | treated | treated |
| Group 3 (T3) | control | control | treated |
| Group 4 (T4) | Control | control | control |

A phased-in RCT will be possible when the delivery of interventions does not begin at the same time in all target communities and the timing of interventions can be randomized. Such randomization will mitigate potential selection bias that may arise if, for example, motivated communities or those with political/economic influence are more likely to receive treatment first. For this approach to be feasible, however, all communities to be targeted in a phase-in manner should be known beforehand (both by implementers and evaluators) to collect baseline data.

**Baseline survey design**: While a baseline data can be used both for RO1 and impact assessment, the type and scale of the baseline survey will be determined by the evaluation approach to be pursued. If no RCTs are conducted, for example, baseline data may not need to

[3] Isolation of control sites is necessary to limit potential contamination effect and ensure that control sites can serve as a valid counterfactual.

[4] The idea of this type of evaluation is to use, as the comparison group, people who have applied for a program but not yet received it. Random and non-experimental pipeline comparisons have also been used. It is sometimes called "pipeline matching" in the literature

be collected from non-beneficiary communities/households. In addition, baseline survey instrument does not need to be as detailed as the draft (LSMS-type) instrument IFPRI and the research teams discussed and put together, to gather data on a wide variety of topics such as agricultural practices, inputs, outputs, health, labor, shocks, poverty, nutrition, and income diversification.

**Survey timing:** Ideally, the baseline survey needs to be collected before implementation starts. Nevertheless, since the main objective of AR is to "to provide pathways out of hunger and poverty for small holder families through sustainably intensified farming systems", this means that one of the principal aspects to look at is how the farmer manages the inputs to obtain the outputs, and hence a careful measurement of both should be sought. It was decided to look at the change in production per hectare as the main output variable, and this means that the timing of the baseline surveys needs to be set in order to minimize the measurement error due to recall bias in agricultural production. The table below summarizes the timeline for baseline survey.

| Country | Fielding Date | Survey Duration | Time of data availability | Remarks |
|---|---|---|---|---|
| Northern Ghana | March/April, 2013 | One month | June/July, 2013 | |
| Mali | TBD | TBD | TBD | |
| Tanzania | June/July, 2013 | One month | September/October | |
| Malawi | Summer, 2013 | One month | TBD | |
| Ethiopia | October/November, 2013 | One month | January/February, 2014 | Data needed for RO1 soon |

**Power calculation** - While an RCT is considered as the gold standard for rigorous impact assessment, even the most rigorously executed RCTs may fail to correctly answer the research question on the quantitative impact if the sample size is too small to detect a minimum effect size of the intervention. When planning an RCT evaluation design, therefore, one needs to conduct sample size estimation to calculate an appropriate sample size, for a given study design, to simultaneously achieve the desired statistical significance level and power.[5] Desired sample size is calculated based on assumptions about:

- Statistical significance level (usually set at 5% or 1%),
- Power of the design (usually set at 80% or 90%),
- Effect size for the outcome of interest (to be determined by the project),
- Intra-cluster correlation of the outcome variable (to be estimated from previous studies or samples), and
- Outcome variable correlation between baseline and follow-up measurements.

To determine the minimal data (or sample size) required to detect a significant research finding, IFPRI conducted initial sample size calculations. Below are the results of IFPRI's power calculation for Ghana, Tanzania, Malawi, and Ethiopia, along with survey budget estimates under different scenarios of data requirements.

### I. West Africa mega-site[6]

**Northern Ghana** - IFPRI conducted sample size calculation for Northern Ghana using agricultural household data from the 5th round Ghana using Living Standards Survey (GLSS5) conducted in 2005/06. The GLSS5 employed a two-stage stratified cluster sampling (agro-ecological zone and rural/urban as stratum, Enumeration Area as a first stage of selection and household as the second stage) and collected detailed socioeconomic and demographic data from 8,687 nationally representative households residing in 580 enumeration areas across the country. A subset of the GLSS5 data (for Upper West, Upper East and Northern Region) is used for sample size calculation.

Although AR activities are expected to have system-level effect, Maize harvest value per hectare (MHVpH) is used to calculate desired sample size. AR implementers in Northern Ghana expect a 20% increase in MHVpH between baseline and follow-up as a result of AR research

[5] Statistical significance level refer to the probability that the research design will reject the null hypothesis of "no impact" when in fact the null is true; while statistical power refers to the probability that the design will reject the null hypothesis of "no impact" when the null is false.
[6] IFPRI will conduct sample size calculations for Northern Mali.

activities. A correlation of 0.7 is assumed in MHVpH between baseline and follow-up. The table below summarizes results from sample size calculation under alternative assumptions on statistical power, intra-cluster correlation, number of sample farmers per community, and estimated cost of survey data collection per household.

Sample Size Calculation and Budget Estimates for Ghana

| | **Follow-up scenarios -> 20% increase in avg. maize harvest value/area** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | New maize harvest/ha | Correlation between measurements | Power | ρ[7] | Sample required (N) | # of households/village | # of villages | Cost of the survey 2013 ('000)[8] | | |
| | | | | | | | | $80 | $100 | $150 |
| **Baseline values:**<br><br>Avg maize harvest value/ha: 192 GHc/ha<br><br>Std. dev.: 401<br><br>Deff[9]: 3.41 (ρ=.172) | 230 | 0.7 | 90% | - | 1,125 | - | - | $90 | $113 | $169 |
| | | | | 0.172 | 2,867 | 10 | 287 | $229 | $287 | $430 |
| | | | | 0.1 | 2,138 | 10 | 214 | $171 | $214 | $321 |
| | | | | 0.05 | 1,632 | 10 | 163 | $131 | $163 | $245 |
| | | | | 0.03 | 1,429 | 10 | 143 | $114 | $143 | $214 |
| | | | 80% | - | 812 | - | - | $65 | $81 | $122 |
| | | | | 0.172 | 2,070 | 10 | 207 | $166 | $207 | $311 |
| | | | | 0.1 | 1,543 | 10 | 154 | $123 | $154 | $231 |
| | | | | 0.05 | 1,178 | 10 | 118 | $94 | $118 | $177 |
| | | | | 0.03 | 1,032 | 10 | 103 | $83 | $103 | $155 |
| | | | | 0.172 | 3,467 | 20 | 173 | $277 | $347 | $520 |
| | | | | 0.1 | 2,355 | 20 | 118 | $188 | $236 | $353 |
| | | | | 0.05 | 1,584 | 20 | 79 | $127 | $158 | $238 |
| | | | | 0.03 | 1,275 | 20 | 64 | $102 | $128 | $191 |

[7] ρ stands for intra-cluster correlation.

[8] The last three cost columns summarize survey budget estimate, assuming survey cost per farm household of $80, $100, and $150, respectively. The actual per unit cost will depend on factors such as village's access to transportation, timing of the survey (rainy versus dry season), and the number villages to be surveyed.

[9] "Deff" stands for design effect, a measure of the loss of effectiveness of data due to the use of cluster sampling as opposed to simple random sampling. Deff = 3.4 means that the sample variance under cluster sampling (by Enumeration Area) of GLSS5 is 3.4 times bigger than it would be if the survey were based on the same sample size but selected randomly. Alternatively, it means that only about one-third as many sample cases would be needed to measure the given statistic if a simple random sample were used instead of the cluster sample with its design effect of 3.4.

Source: IFPRI calculation based on data from GLSS5 (2005) for the Upper East, Upper West, and Northern Regions.

For example, assuming an 80% chance that our RCT design will correctly conclude a significant effect when it really exist in the population, 1,584 households are needed with $\rho$ equals 0.05 and 1,275 with $\rho$ equals 0.03. These are minimum sample sizes desired and are not adjusted for non-response and attrition between baseline and follow-up. Adjusting the sample size upwards by 10% may be desirable to give further cushion for the likelihood of non-response, or to further boost the power of the design, if the sample size is achieved.[10]

Assuming 20 sample farmers per community[11], for example, a minimum of 64 or 80 communities (32/40 beneficiary and 32/40 control communities) would be needed, depending on the assumed $\rho$, to achieve an 80% chance that the RCT design will identify program impact when in fact there is one. As discussed in detail in the site selection report for northern Ghana, 52 communities (excluding Tamale) have been proposed from 5 development domains stratified by the length of growth period and market access.[12]

In February 2013, a team of IITA-Ghana and personnel from the Ministry of Food and Agriculture visited 61 potential intervention communities – 12 in the Upper East, 16 in the Upper West, and 33 in the Northern regions – and gathered information about community size[13], agricultural potential, accessibility during the rainy season, main cropping system, and type of rice farming (rain-fed versus irrigated). The team selected 25 communities (5 in the Upper East, 10 in the Upper West, and 5 in the Northern regions) to be targeted by Africa RISING activities.

---

[10] Designing an experiment with a stronger power (90%) of detecting a real effect would require increasing the sample size, all else equal. For example, we would need 1,632 households (as opposed to 1,584 households with 80% power) with $\rho = 0.05$.

[11] According to a report on Feed the Future Ghana baseline survey (2012) , for example, Enumeration Areas in Ghana tend to be very homogeneous with respect to many key characteristics and taking larger sample within an EA would not add much information. This in turn implies that increasing the number of sample villages is more important than increasing the number of sample farmers per village to increase data efficiency, given the higher likelihood that farmers' characteristics will be more diverse across villages than within village

[12] Refer to the site selection report for stratification and characterization of the proposed communities.

[13] Three classes of communities have been identified based on the number of hamlets (small if 1-50 hamlets, medium if 51-100 hamlets, and large if more than100 hamlets.

## II. East and Southern Africa Mega-site[14]

**Tanzania** - IFPRI conducted sample calculation for Tanzania using household data from Tanzanian Agricultural Sample Census survey that was conducted in 2007/08. The sample calculation is assuming a 10% increase in average maize yield as a result of AR research activities and a 0.7 correlation in maize yield between baseline and follow-up. The sample size calculation is for Babati, Kongwa, and Kiteto (and does not include Mvomero, a potential substitute district for Kilomero)

Sample Size Calculation and Budget Estimates for Tanzania (Babati, Kongwa, and Kiteto)

| | **Follow-up scenarios -> 10% increase in avg. maize yield** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | New maize yield | Correlation between measurements | Power | P | Sample required (N) | # of households/ village | # of villages | Cost of the survey 2013 ('000) | | |
| | | | | | | | | $80 | $100 | $150 |
| **Baseline values:**<br>Avg maize yield: 1660kg/ha<br>Std. dev.: 1311<br>Deff: 4.28 (ρ=.234) | 1826 | 0.7 | 90% | - | 642 | - | - | $51 | $64 | $96 |
| | | | | 0.234 | 1,995 | 10 | 200 | $160 | $200 | $299 |
| | | | | 0.1 | 1,220 | 10 | 122 | $98 | $122 | $183 |
| | | | | 0.05 | 931 | 10 | 93 | $74 | $93 | $140 |
| | | | | 0.03 | 816 | 10 | 82 | $65 | $82 | $122 |
| | | | 80% | - | 464 | - | - | $37 | $46 | $70 |
| | | | | 0.234 | 1442 | 10 | 144 | $115 | $144 | $216 |
| | | | | 0.1 | 882 | 10 | 88 | $71 | $88 | $132 |
| | | | | 0.05 | 673 | 10 | 67 | $54 | $67 | $101 |
| | | | | 0.03 | 590 | 10 | 59 | $47 | $59 | $89 |
| | | | | 0.234 | 2,527 | 20 | 126 | $202 | $253 | $379 |
| | | | | 0.1 | 1,346 | 20 | 67 | $108 | $135 | $202 |
| | | | | 0.05 | 905 | 20 | 45 | $72 | $91 | $136 |
| | | | | 0.03 | 729 | 20 | 36 | $58 | $73 | $109 |

Source: IFPRI calculation based on data from Tanzanian Agricultural Sample Census survey, 2007/08.

[14] IFPRI will conduct sample size calculations for Zambia.

Assuming 20 sample farmers per village, for example, 36 or 45 villages (18/22 beneficiary and 18/22 control village) would be needed, depending on the assumed $\rho$, to achieve an 80% chance that the RCT design will identify program impact when in fact there is one. As discussed in the project proposals in detail, implementers in Tanzania identified 8 research villages (3 in Babati and 5 in Kongwa-Kiteto) for the 2012-2013 season.[15]

**Malawi** – For Malawi, IFPRI conducted sample calculation based on data from the Malawi Integrated Household Survey (IHS) conducted in 2004/05.[16] The sample calculation is assuming a 20% increase in average maize yield as a result of AR research activities and a 0.7 correlation in maize yield measurement between baseline and follow-up.



[15] The eight villages are Long, Sabilo, and Seloto in Babati district; Chitego, Laikala, Mlali, Moleti in Kongwa distrct; and finally Mvugala in Kiteto district.

[16] Maize yield data is only for mono-cropper farmers.

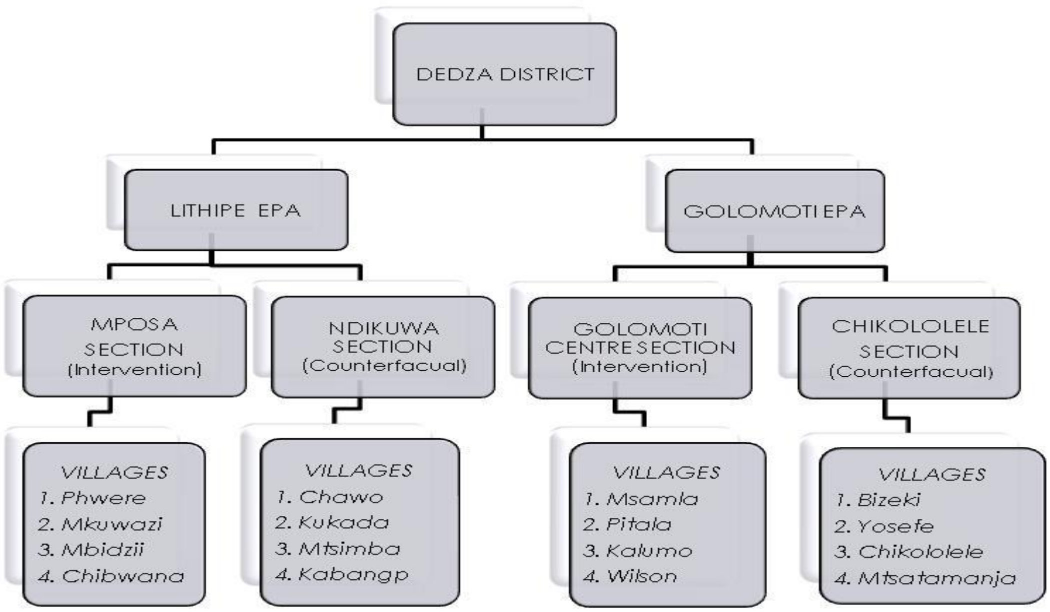
Sample Size Calculation and Budget Estimates for Malawi (Dedza and Ntcheu)

| | **Follow-up scenarios -> 20% increase in avg. maize yield** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | New maize yield | Correlation between measurements | Power | ρ | Sample required (N) | # of households/village | # of villages | Cost of the survey 2013 ('000) | | |
| | | | | | | | | $80 | $100 | $150 |
| **Baseline values:** Avg maize yield: 049kg/ha Std. dev.: 1955 Deff: 2.36 (ρ=.072) | 1259 | 0.7 | 90% | - | 893 | - | - | $71 | $89 | $134 |
| | | | | 0.072 | 1,469 | 10 | 147 | $118 | $147 | $220 |
| | | | | 0.1 | 1,697 | 10 | 170 | $136 | $170 | $255 |
| | | | | 0.05 | 1,295 | 10 | 130 | $104 | $130 | $194 |
| | | | | 0.03 | 1,135 | 10 | 114 | $91 | $114 | $170 |
| | | | 80% | - | 645 | - | - | $52 | $65 | $97 |
| | | | | 0.072 | 1061 | 10 | 106 | $85 | $106 | $159 |
| | | | | 0.1 | 1226 | 10 | 123 | $98 | $123 | $184 |
| | | | | 0.05 | 936 | 10 | 94 | $75 | $94 | $140 |
| | | | | 0.03 | 820 | 10 | 82 | $66 | $82 | $123 |
| | | | | 0.072 | 1,523 | 20 | 76 | $122 | $152 | $228 |
| | | | | 0.1 | 1,871 | 20 | 94 | $150 | $187 | $281 |
| | | | | 0.05 | 1,258 | 20 | 63 | $101 | $126 | $189 |
| | | | | 0.03 | 1,013 | 20 | 51 | $81 | $101 | $152 |

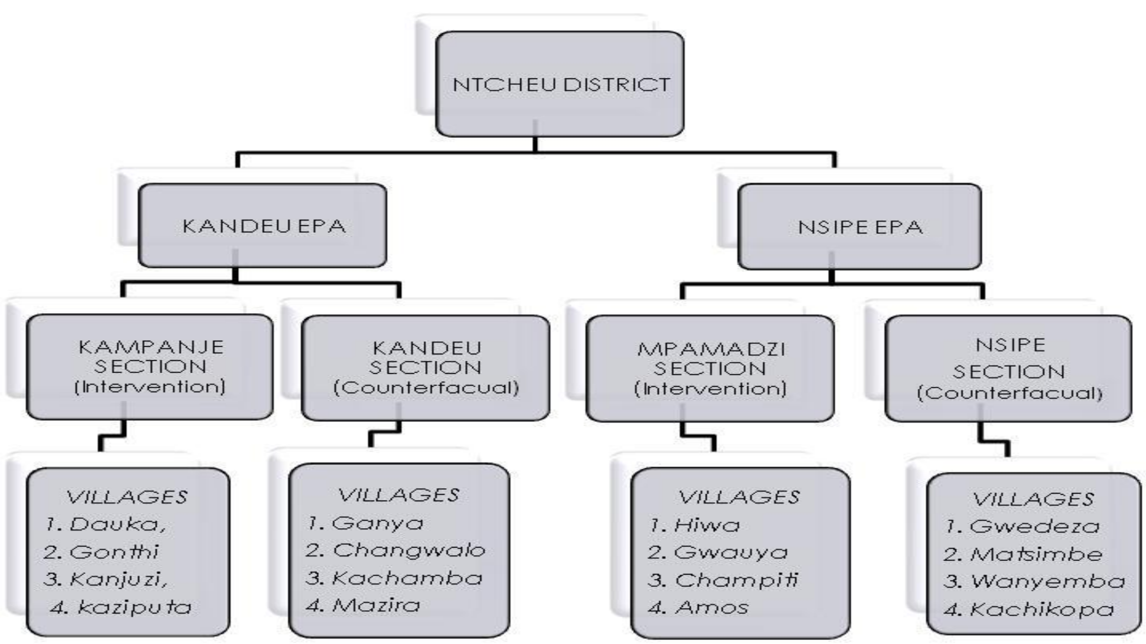
Source: IFPRI calculation based on data from Malawi Integrated Household Survey, 2004/05.

The below organograms show the location of proposed research and control sections and villages by Extension Planning Area (EPA) in Dedza and Ntcheu districts initially selected by Michigan State University (MSU).

Dedza district: Africa RISING Intervention and Counterfactual villages




Ntcheu district: Africa RISING Intervention and Counterfactual villages



Assuming 20 sample farmers per village, for example, 51 or 63 villages (25/31 beneficiary and 25/31 control) would be needed, depending on the assumed $\rho$, to achieve an 80% chance that the RCT design will identify program impact when in fact there is one. MSU is planning to intervene in 16 villages in the first year, expecting to move into the 3rd EPA in each of the districts during the second year. As discussed in the site selection report for Malawi, IFPRI believes that the proposed control sites are too close to research sites (especially for Kandeu EPA, 7km between Kampanje and Kandeu sections) to avoid potential contamination and suggests, if possible, reselection of control sites that are farther away from research sites.

### III. Ethiopian Highlands Mega-site

For Ethiopia, IFPRI conducted sample calculation using data from the Ethiopian Agricultural Survey conducted in 2001. The sample size calculation assumes a 10% increase in average wheat yield as a result of AR research activities and a 0.7 correlation in maize yield measurement between baseline and follow-up.

Sample Size Calculation and Budget Estimates for Ethiopia (Amhara, Oromia, SNNPR, and Tigray)

| | **Follow-up scenarios -> 10% increase in avg. wheat yield** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | New wheat yield | Correlation b/n measurements | Power | ρ | Sample required (N) | # of households/village | # of villages | Cost of the survey 2013 ('000) | | |
| | | | | | | | | $80 | $100 | $150 |
| **Baseline values:** Avg wheat yield: 1171 kg/ha Std. dev.: 554 Deff: 14.15 (ρ=.453) | 1288 | 0.7 | 90% | - | 230 | - | - | $18 | $23 | $35 |
| | | | | 0.453 | 1,169 | 10 | 117 | $94 | $117 | $175 |
| | | | | 0.1 | 437 | 10 | 44 | $35 | $44 | $66 |
| | | | | 0.05 | 334 | 10 | 33 | $27 | $33 | $50 |
| | | | | 0.03 | 293 | 10 | 29 | $23 | $29 | $44 |
| | | | 80% | - | 166 | - | - | $13 | $17 | $25 |
| | | | | 0.453 | 844 | 10 | 84 | $68 | $84 | $127 |
| | | | | 0.1 | 316 | 10 | 32 | $25 | $32 | $47 |
| | | | | 0.05 | 241 | 10 | 24 | $19 | $24 | $36 |
| | | | | 0.03 | 211 | 10 | 21 | $17 | $21 | $32 |
| | | | | 0.453 | 1,596 | 20 | 80 | $128 | $160 | $239 |
| | | | | 0.1 | 482 | 20 | 24 | $39 | $48 | $72 |
| | | | | 0.05 | 324 | 20 | 16 | $26 | $32 | $49 |
| | | | | 0.03 | 261 | 20 | 13 | $21 | $26 | $39 |

Source: IFPRI calculation based on data from the Ethiopian Agricultural Survey conducted, 2001.

Assuming 20 sample farmers per village (*Kebele)*, for example, 13 or 16 communities (6/8 beneficiary and 6/8 control communities) would be needed, depending on the assumed ρ, to achieve an 80% chance that the RCT design will identify program impact when in fact there is one. The Ethiopian Highlands team has identified 8 *Kebeles* for initial engagement (with 30 farmers per village expected to be targeted) with the number of research *Kebeles* to increase to 18-24 in the future.[17]

## 3. Alternative/Complementary approach to RCTs

### 3.1. Farming Systems Modeling

As discussed in a draft proposal put together by AR implementers,[18] farming system modeling would involve baseline data collection (to construct structural farm typologies), rapid farming system characterization (to develop functional farm typologies), detailed characterization of farming system description, and complete diagnosis of the farming system.

The farming systems approach looks at all the inputs at the farm level, the crop-livestock interaction, yield gap, and models a "likely" impact into the future, based on the initial parameters and applying some simulation scenarios. As such, there will not be a need for an extensive and statistically representative data collection to calibrate the model. It would be necessary to collect the relevant information among the farmers beneficiary of the intervention only, without adhering to a strict sampling frame.[19]Data can be collected without concerns about statistical power or representativeness, and this approach would not control for potential confounding factors (i.e., factors that affect both farmers' selection to participate in AR and farmer-level outcomes as a result of AR intervention). In addition, the exercise would be static, disturbed by scenario parameters set arbitrarily by the researcher. While data requirements for traditional farming system models (NUANCES, APSFarm) are lower than for RCTs , results from farming systems modeling would not provide evidence on the causal impact of AR.

### 3.2. Mother - baby trials

[17] The eight Kebeles are Salka and Ilu Sanbitu (in Sinana woreda in Oromia region), Jawe and Upper Gana (in Lemo woreda in SNNP region), Embahasti and Tsibet (in Endamehoni woreda in Tigri region), and Gudo Beret and Goshe Bado (in Basona-Werana woreda in Amhara region).

[18] Refer to the "Fast-tracking farming system analysis activities within Africa RISING – a proposal – February 2013" for more on farming systems modeling.

19 In the case of different projects targeting farmers with diverse characteristics, as Africa RISING in each mega-site is the set-up of a sampling frame would involve the time consuming task of a farmer household listing tailored to the specific criteria used for targeting in each project.

A 'mother' trial -a researcher-controlled station where multi-treatment experiments are conducted- combined with 'baby' trials -where farmers choose technologies from the mother that better fit their socioeconomic circumstances- is one approach that can is being adopted in the context of AR in Malawi. Such mother-baby trials can be conducted in several sites per village to capture farmers' adoption decision under different biophysical and socio-economic conditions. While results from such approach will generate valuable information about effectiveness of technologies and their diffusion[20], assessing causal impact of research activities on technology adoption and other socioeconomic outcomes may not be feasible if farmers participating in baby trials are inherently different from the farmers in the reference universe (e.g., if the targeted farmers have higher motivation, better skill set, or equipment due to previous participation in similar programs), resulting in selection bias. Such approach could also induce farmers to manage their ("baby") trials differently than they would if they were to adopt the technology independently, that is regardless of their interaction with agricultural extension workers and others who supervise them.

### 3.3. Qualitative Methods

Qualitative methods (e.g., focus group discussions, in-depth interviews) are commonly used in evaluations to explore specific facets of programs and to give voice to participants' experiences. These methods cannot provide evidence on the causal impact of a program but can provide in-depth information to assist in enhancing the quality of program. Findings from such studies "are not thought of as facts that are applicable to the population at large, but rather as descriptions, notions, or theories applicable within a specified setting" Malterud (2001, 486).

Qualitative data analyses bring numbers and stories (about causal mechanism) together and, when used to expand current program evaluation, can provide a powerful tool to help programs assess their progress, identify areas needing improvement, and help families, programs and funding agencies recognize and gauge their successes. Difficulties with external validity to the wider population, replication (and ability to independently verifying results),

[20] Diffusion of technologies can be assessed within the framework of social learning (from neighbors, friends, and relatives (Foster and Rosenzweig, 2010 among others).

subjectivity in interpretation of results, and not-quantifiable measurement of outcomes are among the challenges associated with qualitative evaluation methods.

### 3.4. Quasi-experimental approaches

Quasi-experimental methods such as matching[21] and regression discontinuity[22] seek to solve the 'evaluation problem' by using data on program non-participants, or on participants at a different time, as the basis for estimating outcomes that would have occurred for participants in the absence of program participation. They try to recover the information on the participants had not been treated, reconstructing the "missing counterfactuals". Quasi-experimental approaches use econometric and statistical methods to control for bias due to observable selection (as selection on unobservables can be fulfilled only through RCTs) and may be appropriate when:

- Random assignment is not feasible, (perceived to be) unethical, or unaffordable,
- When the program is still under development and there is a need to wait until the program is well-developed and has settled into a clear and consistent set of activities,
- If the pool of potential participants is too small to fill both treatment and control groups,
- If it is impossible to avoid "contamination" of the control group, or
- If the program wishes to establish ongoing, internal evaluation capacity

Some of the risks in adopting a quasi-experimental approach lie on the selection of a comparison population different from the beneficiary population[23] and the high cost of data collection for both the target and the control population before and after the program. If the latter case, it may be worth considering whether an experimental study might bear a more

[21] The matching method estimates the program impact by comparing outcomes for program participants and non-participants in the time period(s) after the program commences. This method uses data on outcomes of non-participants in the period after program commencement to estimate non-participation outcomes for the group of participants.

[22] A regression discontinuity method estimates the program impact by comparing outcomes for program participants and non-participants who are, respectively, 'just above' and 'just below' the threshold level of some characteristic that defines eligibility for participation. Such approach can be employed when program participation is a deterministic and scontinuous function of some observable characteristic (such as total expenditure, land holding).

[23] For example, if the comparison population is more advantaged than the population being served, then outcomes for program participants may seem less positive than they really are. Alternatively, if something happens to the comparison population – for example, if it gets served by another development program (e.g., NAFAKA in Tanzania or ATA in Ethiopia) then the value of the comparison will be undermined.

cost-effective design, since such design would allow statements on whether the program had a statistically significant impact on outcomes for participants.

Quasi-experimental studies can inform discussions of cause and effect but, unlike randomized experiments, they cannot definitively establish a causal link, given their selection on observable characteristcs only. Nonetheless, these approaches can be very valuable in providing descriptive information about the population served, whether anticipated changes are occurring for some groups, the magnitude of change that is occurring over time, , and whether some outcomes are changing while others are not.

## 4. Concluding remark

Although the use of RCTs to estimate the causal impact of AR will increase the resources needed and the amount of work, this approach should be considered seriously (at least in some of the focus countries or research activities in the coming years) to rigorously evaluate the impact of AR activities. RCTs should be complemented with other approaches (e.g., farming system modeling, qualitative analysis) to generate more robust evidence on causal impact and underlying mechanisms. As discussed in detail in Section 2, however, designing an RCT is subject to a number of conditions that, if not fulfilled, would require the AR team to explore other options outlined in Section 3. IFPRI and implementers would therefore need to carefully assess the type, nature, and timing of planned activities and target households to determine if RCTs can be imbedded into the specific projects, and agree upon other alternative/complementary evaluation approaches that are appropriate to the specific characteristics of research activities (to be) conducted.