International Initiative for Impact Evaluation
Improving lives through impact evaluation

Pre-analysis plan (PAP)

Evaluating Agricultural Information Creation and Dissemination in Western Kenya **TW4.1011**

NOTE (April 2015): This pre-analysis plan should be updated once the partner organization provides more details about the information content of one of their intervention arms which is currently being finalized (and before the endline is conducted). We expect to do this, before the next reporting period.

1. Purpose

1.1. Pre-analysis Plans (PAPs)

PAPs are a tool for planning and pre-committing to the main outcomes and treatments of interest in a study as well as the specifications for estimating impacts. While the authors may explore additional hypotheses, outcomes and specifications, these should be reported as having been developed after drafting the pre-analysis plan and after engaging with the data.

1.2. For 3ie

The PAP will build on the revised technical proposal and become a key reference document for assessing the completeness of the Final Report.

2. Timing

Ideally, the plan will be written following the baseline data collection and prior to assembling the next follow-up survey instruments. This should be reflected in the D&D.

3. Methods overview

This document provides a pre-analysis plan for the evaluation of several interventions that provided agricultural information to smallholder farmers in Western Kenya. The implementation of these interventions was done in coordination with the Kenya Agriculture and Livestock Research Organization (KALRO). The data was collected by Innovations for Poverty Action (IPA) Kenya.

The document is outlined as follows: we first describe the interventions. We then discuss the data sources, identification strategy, treatment assignment and the way in which the sample was selected. We then outline hypotheses and the empirical strategies to test them.

3.1. Interventions

This project seeks to measure the effects on beliefs, knowledge and input use of providing locally relevant agricultural information to farmers. We evaluate different information delivery channels:

(a) Farmer Open Days with Demonstration Plots

Traditionally, one of KALRO's main strategies to provide information to farmers about best agricultural practices is to organize large events around farming demonstration sites, where they showcase appropriate agricultural technologies for the respective area agro-ecologies. Through different outreach activities, farmers living around the demonstration sites are invited to attend and talk to different agricultural experts. The goal of these events is to serve as learning platform for farmers.

(b) E-extension Activities

Starting in 2015, KALRO plans to provide agricultural recommendations to farmers through mobile phones. The aim of the messaging system is to bridge problems associated with access and knowledge gaps by directly connecting farmers with agricultural information and other market players.

To implement this activity, KALRO is partnering with the Ministry of Agriculture nationwide e-extension program. KALRO is currently working on setting up the systems to provide these services in the study sites.

(c) Intensive Information Delivery

Research work in other contexts has found that farmers might fail to learn about profitable agricultural practices because of failures to notice important dimensions of the production process (Hanna, Mullainathan and Schwartzstein, 2014). In order to understand how much farmers learn from actively experimenting in their own farms (rather than receiving information about local means), researchers worked with an additional sample of farmers to set up on-farm experimental plots. This will help identify potential constraints and opportunities in the learning process.

3.2. Identification Strategy and Treatment Assignment

We evaluate the effectiveness of these interventions via randomized control trials. In order to evaluate the impacts of the Farmer Open Days events and the e-extension program, we identified a group of farmers who were eligible to participate in these programs (using KALRO's criteria) and completed a baseline survey with them in mid-2014 (the details of sample selection are discussed below). Once the baseline survey was completed we randomized farmers into three separate groups: (1) assigned to farmer open days, (2) assigned to the e-extension program and (3) assigned to a comparison group.

The first round of open field days were conducted in late 2014. All farmers randomized into the open field day events were invited to attend. Since we wanted to evaluate the effects of attending the events on subsequent knowledge and input use, we facilitated attendance by offering transport to the sites. Attendance information was collected at these events.

The SMS intervention will be conducted in 2015. The randomly selected sample of farmers assigned to this treatment will receive text messages with agricultural recommendations.

We will assess changes in beliefs and knowledge about agricultural inputs and practices, as well as, changes in input use, through an endline survey that will be completed in late 2015.

To evaluate the impact of on-farm experimental plots on knowledge and input use, we exploit the fact that the set-up of experimental plots with farmers had been done through a lottery. Therefore we can compare the outcomes of those farmers who received experimental plots in their farms, against the outcomes of those farmers who were not assigned to receive these plots. A harvest survey that contained questions on agricultural knowledge and beliefs, was collected in late 2014. We have not yet analyzed this data.

3.3. Sample selection

In coordination with KALRO, we first selected two subcounties to undertake the evaluation work (Ugenya and Mumias). These subcounties were chosen because KALRO was already planning to undertake various activities in those locations, and they were the most representative of other areas of operation. Once these areas were selected, a team of enumerators conducted census surveys in these areas. Within each subcounty, we first identified primary schools and the sites of KALRO's demonstration to serve as central landmarks. From these sites enumerators used a random walk method to sample respondents. In the random walk method, households are selected to participate in the survey, using a specific walking rule to ensure sample representativeness. Field officers attempted to complete approximately 1,400 census surveys. Subsequently a subsample of individuals from the census sample was selected to be invited to participate in the study. The criteria of inclusion were: (i) respondent or another household member had a phone, (ii) respondent had grown maize or legumes during the previous year and, (iii) respondent were in charge of farming activities for the household. These criteria were used to ensure that the sample would be representative of those farmers who are usually targeted by KARLO.

The sample of farmers who participated in the on-farm experimental plots had been previously selected through a random walk method in Busia subcounty. In order to minimize selection bias, potential respondents were asked whether they would be willing to set up experimental test plots on their land. Only those who were willing to set up these test plots were selected into the sample, and only then they were randomly assigned into receiving this treatment or remaining as part of the comparison group.

3.4. Data sources

The data used in this study will be collected through fieldwork led by Innovations for Poverty Action (IPA) Kenya. There are two main sources of information: self-reported questionnaires conducted at respondents' homes and administrative data from the redemption of input coupons.

(a) Open Farmer days and SMS intervention

In total 1,250 eligible farmers were tracked and completed the baseline survey. The baseline questionnaire was designed to contain information relevant to KALRO's project and the experimental evaluation led by IPA. The questionnaires included questions ondemographics, farming experiences, beliefs and knowledge about agricultural practices. Each survey took approximately 2 hours to complete.

We plan to complete a home-based endline questionnaire at the end of the intervention period, through which we will learn about farmers' knowledge and beliefs, whether information transfer affected the likelihood of using locally adequate inputs, and the adoption of other ISFM practices promoted by KALRO. In addition, in order to measure the impact of these different interventions,

we will use a coupon-redemption technique. At endline we will provide all respondents with coupons (with serial numbers) redeemable for either agricultural inputs (fertilizer and/or lime) or soap at local agro-dealers. A key outcome variable will be whether receiving information increases farmers' coupon redemption for the recommended inputs. In addition, we will compare redemption rates for different information delivery channels. Using coupons will allow the research team to obtain data on actual agricultural input choices made by participants rather than only relying on self-reported information.

(b) On-farm experimental plots

In order to evaluate the effects of the on-farm experimental plots, IPA conducted a baseline and endline survey with all farmers in the treatment and control group in 2014. The surveys contained questions about knowledge and beliefs about the effectiveness of different types of agricultural inputs. Coupons redeemable for fertilizer were provided to everyone in this sample.

4. Hypotheses under investigation

We are collecting several indicators to test the effects and mechanisms of behavior change (if any) of exposure to treatment. The hypotheses can be grouped into different categories:

Category	***Hypotheses***	***Outcome Measurement***
Knowledge	*The interventions may have positive average impacts on farmer's knowledge about existence and appropriateness of agricultural inputs for their land.*	• Questions about awareness of chemical fertilizer and lime • Questions about reasons to apply lime • Questions about knowledge of other agricultural practices • Questions about correct quantities of inputs to apply • Questions about information diffusion.
Beliefs	*The interventions may have positive average impacts on farmer's knowledge about the existence and appropriateness of different agricultural inputs for their land.*	• Questions about perceptions of effectiveness of inputs. • Questions about perceptions of profitability of inputs.
Agricultural Practices	*The interventions may have positive effects on the use of recommended agricultural practices.*	• Questions about adoption of ISFM practices
Agricultural Input Use	*The interventions may have positive average effects on the use of recommended inputs and technologies such as soil testing, chemical fertilizer and lime.*	• Questions about use of inputs (fertilizer and lime) • Coupon redemption questions.

5. Hypotheses, hypothesis testing, and heterogeneities of interest

5.1 General set up

The general empirical strategy will be to run the following regression:

$$y_{i,t=endline} = \beta_0 + \beta_1 Treatment_i + \beta_2 y_{i,t=baseline} + \delta X_{i,t=baseline} + \epsilon_i$$

Where $y_{i,t=endline}$ indicates the post treatment outcomes of interest, *Treatment* represents a dummy variable that takes value 1 to indicate different treatment interventions. The coefficient provides the intent to treat estimates for the treatment. We also control for baseline outcomes when these variables were collected. We also plan to run additional specifications with interaction of treatment and selected baseline characteristics (as described in the next subsection) to identify heterogeneous treatment effects. In order to adjust standard errors for constraints we imposed during the stratified randomization, we plan to include dummies for the different stratum used in the randomization (*X's*). We will use robust standard errors.

Since not all invited farmers will participate in the open farmer days and the e-extension activities. We will also estimate a treatment on the treated specification, in which we will instrument participation with the treatment assignment.

$$y_{i,t=endline} = \beta_0 + \beta_1 Participation_i + \beta_2 y_{i,t=baseline} + \delta X_{i,t=baseline} + \epsilon_i$$

5.2 Hypotheses

Hypothesis 1: Receiving information may improve individual's knowledge about the existence and appropriateness of agricultural practices and inputs.

- We will explore knowledge about new inputs of interest (NPK, CAN and lime) [Baseline: F28, F34, F39, F44] and construct a demeaned index of other inputs and practices [Baseline: G9].

Note: KALRO is still finalizing the content of some of the information provision they will provide and this might affect the choice of knowledge questions used in the endline.

- The underlying assumption in the theory of change is that farmers did not know about these technologies and practices, and the information provision will resolve this knowledge gap. The assumption is that they can understand and remember the information that is provided to them. For the cellphone intervention this also requires certain level of literacy.
- The null hypothesis is that the information will fail to increase the knowledge about the existence of different agricultural practices and inputs. This implies the use of one-sided significance test.
- Since the randomization was conducted at the individual level standard errors will not be clustered.

Hypothesis 2: Receiving information may improve individual's beliefs about profitability and effectiveness of agricultural practices and inputs. (on-farm test plots sample – conditional on results from test plots).

- We will explore beliefs about effectiveness of all key inputs relative to not using them [Baseline: G1-G4, G5, G6].
- Construct a demeaned index of effectiveness of other practices.
- The theory of change, assumes that even if people have (or receive) the knowledge about the new technologies, they also require changing beliefs about the profitability or effectiveness of locally appropriate inputs in order to adopt them. For the experimental, test plots these are changed through the profitability information.
- The effects will depend on the original beliefs about effectiveness and profitability of different types of inputs and practices and the exact information provided to farmers. [April 1, 2015 – Needs to be aligned with content information intervention, which needs to be finalized]
- We will use a two-sided test. The randomization was conducted at the individual level so standard errors will not be clustered.

Hypothesis 3: Receiving new information may spur increased agricultural information diffusion with social networks.

- We will explore the amount of interactions with social networks that occur after the open field day and e-extension treatments [Baseline: I9,10]
- One reason why we have not found much information diffusion in the past is that farmers do not share new information because they don't think there is much to gain from talking to others/there is few information about technologies. The introduction of new information can spur diffusion.
- For experimental test plot sample, we will sample a number of their social networks and measure their knowledge and WTP for information. Standard errors will be clustered at the original respondent level.

Hypothesis 4: Receiving information may improve individuals' adoption of practices and new inputs.

- The outcome indicators for practices will be an index. [Baseline: G10 for e-extension and open field days]
- The outcome indicators for adoption of inputs will contain questions on previous season use of NPK and lime.
- The outcome indicators for adoption of inputs will also use the redemption of coupons at shops.

5.3 Heterogeneous treatment effects and subgroup analyses

We will explore heterogeneity of treatment according to the following variables:

- Baseline knowledge about existence of inputs (NPK and Lime)
- Baseline experience of practices (tercile score for practices index)
- Gender
- Literacy

- Baseline cognitive scores (tercile score for index)
- Baseline beliefs about profitability of inputs
- Baseline beliefs about effectiveness of inputs
- Variable indicating local recommendation. For experimental test plots treatment, rates of return for on-farm plots (ie. input with highest ror for given area)
- Land size

5.4 Hypothesis testing

In order to address issues of multiple hypothesis testing we plan to group the list of practices and inputs that could change through the information intervention into different categories (e.g. Baseline G9 and G10). We will sign the outcomes and then standardize treatment effects as Kling, Katz and Liebman. Second, following Fink et al 2013, we plan to conduct Bonferroni FWER corrections.

6. Addressing less-than-complete data

(a) Survey attrition

Attrition is a first-order concern for any evaluation since it can seriously bias the estimates. Our primary approach to limit this problem will be to intensively track and re-survey all baseline respondents. Researchers and IPAK have extensive institutional knowledge on designing and managing data collection with farmers in Western Kenya and have been able to keep attrition rates below 5% in similar surveys.

In order to check whether there is selective attrition we will first check if attrition levels are correlated with the treatment status. In order to check for selective attrition, we expect to regress an indicator of attrition (either not found or declines to survey) on treatment status. If attrition is found to be statistically significant at the 5 percent level then we will employ Lee bounds to correct for this.

(b) Addressing item non-response (data missingness)

We will estimate whether attrition is related to treatment status by regressing an indicator of attrition on treatment status (with basic controls as explained above). If this is significant, we will again employ Lee bounds. We do not plan to impute data if it is missing.