abt_assoc_logo_pms_cmykAgResults Kenya On-Farm Storage Pilot Evaluation Analysis Plan

March 2015

Prepared for:

**Department for International Development (DFID)**

22 Whitehall

London

SW1A 2EG

Submitted by:

**Abt Associates**

4550 Montgomery Avenue  
Suite 800 North   
Bethesda, MD 20814

**In Partnership with:**

Denise Mainville Consulting

Rockville, MD

# Preface

AgResults is a $110 million multilateral initiative incentivizing and rewarding high-impact agricultural innovations that promote global food security, health, and nutrition through the design and implemen­ta­tion of pull mechanism pilots. It is funded by the governments of Australia, Canada, the United Kingdom, the United States, and the Bill & Melinda Gates Foundation, and managed through a Financial Intermediary Fund operated by the World Bank. By using pull mechanisms, AgResults extends beyond traditional aid measures to promote the adoption of innovative technologies with high-yield development impact. AgResults will provide economic incentives to private sector actors in smallholder agriculture to develop and ensure the uptake of innovative technologies with the potential to yield high development impacts. It will help overcome market failures impeding the establishment of sustainable commercial markets for such technologies, or goods produced by means of them, and thereby achieve substantial and sustained development impacts, manifested in improved food security and food safety, increased smallholder incomes, and better health and nutrition. It will call upon the ingenuity and drive of the private sector to identify and execute the most effective and efficient strategies to achieve development outcomes.

The AgResults program team comprises a Steering Committee, a Secretariat, a Trustee, country-specific pilot implementers, and an external evaluator. The Steering Committee oversees the implementation of AgResults and is comprised of the five donor agencies and the Trustee. The Steering Committee is responsible for strategic oversight of the initiative, including endorsement of key management decisions, approval of concepts and business plans for proposed pilots, and the monitoring of pilots and the initiative as a whole. The Secretariat is responsible for implementation of the AgResults Initiative and reports to the Steering Committee. In order to fulfil its role effectively, the Secretariat develops a close working relationship with the Trustee and ongoing external evaluator. Core functions include appointing and managing pilot implementation and verification agents, sourcing new pilots, and communicating results. As Trustee for AgResults, the World Bank provides an agreed set of financial intermediary services that include receiving funds, holding funds, investing funds, and transferring them to recipients or other agencies for implementation as directed by the Secretariat on behalf of the Steering Committee.

In Kenya, AgResults is funding an on-farm storage pilot that stimulates improved food security through the widespread adoption of improved on-farm grain storage solutions for smallholder farmers in the Rift Valley and Eastern regions of Kenya. The pilot is designed to demonstrate a successful model for developing and marketing low-cost storage solutions for smallholder farmers by offering cash prize payments to private sector companies based on the volume of low-cost storage capacity sold within a given timeframe. The Kenya pilot is scheduled to begin in May 2015 and run for three years.

Agribusiness Systems International (ASI), a wholly controlled affiliate of ACDI/VOCA, serves as Pilot Manager for the Kenya pilot. As Pilot Manager, ASI is managing overall implementation of the pilot. This includes marketing, promoting, and overseeing the selection process of potential implementers; liaising with international organizations and relevant government ministries; reporting on the progress of the pilot and the approved indicators; overseeing verification and managing the final data analysis to calculate funds disbursements; and documenting and reporting lessons learnt to the Secretariat in a Quarterly Report and in a report following an annual lessons learned event with implementers, once the sales period begins. The Secretariat will share lessons learnt with the broader stakeholder community via the AgResults website, blog and social media.

The Steering Committee appointed [Abt Associates Inc.](http://www.abtassociates.com/) to serve as the External Impact Evaluator for the AgResults pilots, including the Kenya on-farm storage pilot. Abt’s role is to determine on a rigorous scientific basis if the pull mechanisms achieve their objectives of catalysing private sector investments that address underlying market failures, resulting in social outcomes that are different from, and better than, what would have happened in the absence of the mechanism introduced by the pilot initiatives. We will also address the sustainability of the results produced in the private market once the pilot incentives are removed. In our role as the External Impact Evaluator, Abt defines an overall evaluation framework for the AgResults Initiative and an impact analysis strategy for each pilot. We implement and analyse field surveys based on established best practices, conduct qualitative market analyses, and communicate evaluation findings to the Steering Committee and wider audiences as needed. Our role is vital to the AgResults learning agenda of understanding the potential of private sector involvement in the development and spread of agricultural innovation. The evaluation timeframe is 2013 to 2019 (assuming extension of the contract following delays in the start of the Kenya pilot).

This document is an abridged version of Abt’s evaluation design report. It presents the overall evaluation framework and methodological approach, the identification strategy for estimating impact at the small farmer household level, model specifications, power analysis, and our analysis plan.

The Kenya evaluation team is led by Tulika Narayan, with the core quantitative team comprising Betsy Ness-Edelstein (also the country coordinator), Sung-Woo Cho, and Fatih Unlu. Denise Mainville from Denise Mainville Consulting leads the qualitative analysis working with Tabitha Nduku, in-country agricultural economist, and the quantitative team, who also contributed to qualitative investigation. Stephen Bell, the AgResults Team Leader, provides technical advice and quality control of research reports. To date, the Kenya pilot’s evaluation design has benefited from comments and inputs from the external peer reviewers Kelsey Jack, Mushfiq Mobarak, and Ephraim Nkonya. The evaluation design report also benefited from comments from the AgResults Steering Committee and members of the Kenya pilot implementation teams at Deloitte and ASI.

# Evaluation Framework and Methods

This introduction summarizes the evaluation framework, research questions, and methods the Abt Associates team is employing to evaluate the Kenya on-farm storage (OFS) pilot. Although the overall evaluation framework is the same for the two regions of pilot implementation, there are important differences in the incentive structures, implementation context, and expected impact (market penetration rates) for the Rift Valley and Eastern regions. Accordingly, much of our analysis will disaggregate results across the two regions.

The Abt evaluation will assess whether the program has met its objectives and be guided by six out of the seven evaluation questions in the overall evaluation framework established by the research protocol (Evaluation Question 4 on demand for derivative products, part of the overall AgResults research agenda, is not relevant for this pilot):

1. What has been the impact of the AgResults pilot on private sector engagement in the development and uptake of on-farm storage?
2. What has been the impact of the AgResults pilot on smallholders’ uptake of on-farm storage?
3. What has been the impact of the AgResults pilot on smallholder income?
4. What is the impact of the AgResults pilot on consumers’ demand for derivative products? *[not relevant for Kenya pilot]*
5. What evidence exists that the effects of the AgResults pilot will be sustainable in the medium to long term?
6. What is the evidence on the scale of any effect on private sector investment and uptake and on the cost-effectiveness of AgResults as an approach?
7. What lessons can be learnt about best practices in the design and implementation of agricultural pull mechanisms?

We will also address, within each of the questions, whether the pilot’s impact is differentiated by gender or poverty status of beneficiaries, and analyse the determinants of any such effects that are identified. We use a mixed-methods approach to address each of the evaluation questions, with either a quantitative or qualitative approach predominating for each particular question and complementary qualitative or quantitative input to enrich the results.

Questions 1, 5, and 6 are related to the pilot’s impact on the value chain and the value chain participants therein. The small numbers of value chain participants and multiple levels of interaction between the participants call for qualitative methods to answer these questions, which will allow for nuanced inquiry into the motivations of market actors, their strategies for engagement in the market, and the aggregate effects of their investments. Therefore, our evaluation assesses the market-level questions on the agenda (questions 1, 5, and 6) using primarily qualitative methods—specifically a structure, conduct, and performance (SCP) framework that is described in more detail in Section 3.1.

We use primarily quantitative (statistics based) methods to assess the pilot’s impact on smallholder awareness, adoption, and use of on-farm storage solutions and subsequent effects on income (proxied by maize revenue) and food security (questions 2 and 3). Specifically, we employ a short interrupted time series (SITS) approach to estimate the impact of the pilot on smallholder farmers by comparing the pre-pilot time trend on key outcomes to post-pilot levels. The SITS design measures the intervention impact as a departure from the expected levels of the outcome measure (in this case smallholders’ uptake of on-farm solutions, and smallholder income) when projected forward in time, as an approximation of what would happen were the treatment not introduced.[[1]](#footnote-1) The SITS approach entails (1) generating a counterfactual for the outcome measure, which represents the expected level of the outcome in the post-intervention period in the absence of the treatment as the projected trend in pre-intervention observations of the outcome measure, and (2) modelling the treatment effect as a deviation of actual post-intervention outcomes from this counterfactual. Our qualitative research enriches and adds nuance to our understanding of the quantitative analyses of questions 2 and 3.

We address Evaluation Questions 5, 6, and 7 by drawing on and synthesizing the results of Evaluation Questions 1, 2, and 3. Table 1 presents these evaluation questions, which have been modified based on the expected impact of the pilot, along with the main method we use to answer the question.

Table . Evaluation questions and approaches

|  |  |  |
| --- | --- | --- |
| **#** | **Evaluation Question** | **Evaluation Method** |
| 1. | What has been the impact of the AgResults pilot on private sector engagement in the development and uptake of on-farm storage? | Theory-based qualitative; SCP focused on the grain value chains for which the on-farm storage is relevant, particularly the maize value chain. |
| 2. | What has been the impact of the AgResults pilot on the uptake of on-farm storage? | Impact evaluation using SITS supplemented by qualitative interviews. |
| 3. | What has been the impact of the AgResults pilot on smallholder’s maize revenue, and food security? | Impact evaluation using SITS supplemented by qualitative interviews for maize revenue, before and after analysis for maize consumption in lean season. |
| 4. | *Not relevant for Kenya pilot.* | *N/A* |
| 5. | What evidence exists that the effects of the AgResults pilot will be sustainable in the medium to the long term? | Synthesis of results from SCP and impact evaluation. |
| 6. | What is the evidence on the scale of any effect on private sector investment and uptake and on the cost-effectiveness of AgResults as an approach? | Synthesis of results from SCP, with focus on market infrastructure and per-unit cost effectiveness of key outcomes from the impact evaluation. |
| 7. | What lessons can be learnt about best practices in the design and implementation of agricultural pull mechanisms? | Compilation of results from all AgResults pilot evaluations. |

# Short Interrupted Time Series Model

The second and third evaluation questions stated above ask about the impact of the AgResults pilot on smallholder actions and outcomes. In particular, these questions pertain to the impact of the pilot on two primary outcome measures: smallholders’ uptake of on-farm storage solutions and smallholders’ maize revenue (which are referred to as “targeted” or “treated” outcome measures throughout this section). In addition, we will also examine the pilot’s impact on smallholder’s food security as measured by number of months the stored grain is available for own consumption (referred to hereon as food security). Furthermore, we plan to also assess the impact of the program on other outcomes such as smallholder’s access to improved storage technologies, and smallholder’s knowledge and attitude about other storage practices that are essential for successful use of the technology. For example, if farmers do not dry the grain properly, or do not clean the grain properly then the improved technologies will not be effective. While the impact of the pilot on these outcomes will be analysed using the same approach that will be used for the two primary outcomes (namely short interrupted time series methodology which is described in more detail below), the statistical power analysis and the determination of the sample size requirements focused on the two primary outcomes.

In what follows, by “impact” we mean the difference between results with the AgResults incentives present in the market—with whatever private sector supplier activities they bring—and what would have happened to the same smallholders without these factors present. Obviously, we cannot obtain data on the latter once the pilot begins in the target areas of Kenya. In this section, we describe how we will answer these questions in light of this challenge. The goal is to attribute technology adoption and income results to the AgResults pilot and other causal factors such as pre-existing market trends and shifts in agricultural inputs of other sorts. We begin with a description of the “short interrupted time series” approach for conducting this analysis, which is intended to pull apart influences on farmer outcomes to isolate the effect of the pilot intervention. Subsequent subsections give the detailed model for our planned statistical analysis and consider statistical power to detect impacts and the required sample sizes in the Rift Valley and Eastern regions.

## Short interrupted time-series design

Randomized control trials (RCTs)—which split potential program participants into statistically equivalent groups through a lottery and provide the intervention to just one of those groups—are regarded as the gold standard for calculating impacts of programs. Due to the randomness of the split, an RCT results in the study’s “treatment” and “control” groups being balanced on both observable and unobservable characteristics. Hence, their outcomes will differ only to the extent that the intervention has an impact on the treatment group; it is this outcome difference that reveals the intervention’s effects.

When RCTs are not feasible, researchers utilize quasi-experimental designs (QEDs), which compare outcomes of program participants (the treatment group) to a counterfactual that represents their hypothetical outcome levels in the absence of the treatment. An RCT is not feasible for the on-farm storage pilot because implementers have a business incentive to market their technologies to all target areas for the pilot and to all farmers in those areas (through broad advertising and other means), leaving no potential for an equivalent “untreated” set of farmers to be drawn from that pool at random.

**Selection of a quasi-experimental design option.** The most commonly used QED entails comparing the outcomes of the treatment group to a group of units that has similar baseline characteristics to the treatment group members but is not exposed to the treatment—the so-called “comparison group”. By minimizing the observable differences between the two groups, this method alleviates the bias in the treatment effect estimates that could arise due to the influence of other programs, interventions, or initiatives that may act as confounders.

The business plan for AgResults proposed to estimate the effects of the pilot program by comparing two outcome measures—smallholders’ uptake of on-farm storage and income—between a treatment group that consists of smallholders in areas targeted by the program and a comparison group that consists of smallholders in similar but untargeted areas. During our visit to the potential target areas in April 2014, we observed that this strategy may not be feasible due to (1) potentially substantial differences between targeted and untargeted areas and (2) the extensive effort that may be required at the onset of the evaluation to identify appropriate farmers to study in comparison sites.

As a result, we turned our attention to a QED that uses multiple pre-intervention data points to form the counterfactual, the short (or abbreviated) interrupted time-series (SITS) design. The SITS design measures the intervention impact as a departure from the expected levels of the outcome measure (in this case smallholders’ uptake of on-farm solutions, smallholder income or food security) were the treatment not introduced (e.g., Shadish, Cook, and Campbell 2002; Bloom 2003). Specifically, this design entails (1) generating a counterfactual for the outcome measure, which represents the expected level of the outcome in the post-intervention period in the absence of the treatment and is constructed by projecting the trend in pre-intervention observations of the outcome measure, and (2) modelling the treatment effect as a deviation from this counterfactual.

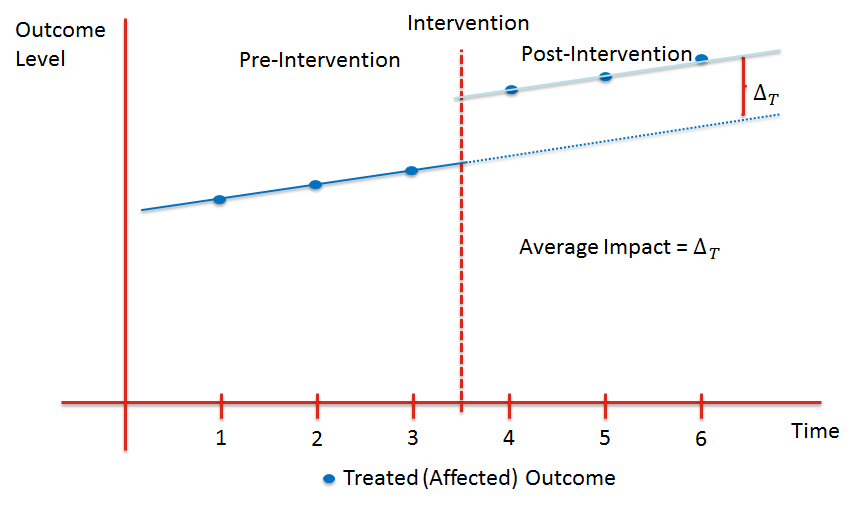
**The short interrupted time-series approach.** Figures 3a and 3b demonstrate a stylized application of the SITS approach for estimating the effect of a treatment on a targeted (or treated) outcome measure with a positive linear trend in the pre-intervention period. In these figures, the y-axis shows the average level of the outcome for the sample, and the x-axis represents the time points. There are three pre-intervention observations in this illustration (time=1, 2, and 3) and three post-intervention observations (time=4, 5, and 6) with the treatment introduced between the third and fourth time points. Pre-intervention time points will be collected through retrospective questions to farmers (on the baseline survey) about previous harvests and practices. The solid line in the pre-intervention period tracks outcomes prior to the pilot and captures the upward trend in the outcome measure. The counterfactual, which is represented by the dotted line, is obtained by extrapolating the pre-program trend information into the post-intervention period following the solid line upward to the right.

Figure 1a. Short interrupted time series with linear trend and year-specific impact estimates



The treatment effect in the first post-intervention period (t=4), represented by , equals the difference between the average outcome level at t=4 and the value of the counterfactual at that time period. Treatment effects at the second and third post-intervention periods are calculated in a similar manner. Alternatively, the average treatment effect across the three post-intervention time periods can be estimated by constructing a line through these time points based on the baseline trend (light blue solid line in Figure 1b) and calculating the difference between the intercept of this line and the intercept of the counterfactual line (denoted by in Figure 3b).

Figure 1b. Short interrupted time series with linear trend and average treatment effect estimate



It is important to note that while a SITS analysis accounts for secular trends in the outcome measure in the estimation of the treatment effect (which is demonstrated in more detail in the next section), it is subject to bias when other factors (i.e. confounders) that influence the outcomes of interest change simultaneously with the arrival of the focal intervention. For example, unusual changes in rainfall or pest burden that coincide with the roll-out of the treatment may be potential confounders as they may also push off their pre-existing trend lines. Unusually heavy rainfall during the first treatment year could increase the maize yield and consequently smallholders’ income, a change that would be inaccurately interpreted as an effect of the pilot unless it is accounted for. Other government or NGO programs that are enacted in the post-treatment period and target the same outcome measures may also act as confounders. If the effect of such confounders on the outcome is in the same direction as the true treatment effect (which is assumed to be positive), not controlling for them would create an upward bias in the estimated effect of the pilot while confounders affecting the outcome in the opposite direction would lead to a downwardly-biased treatment effect estimate. The amount of bias is a direct function of the strength of the confounding influences and may be substantial.

Fortunately, there are at least two ways the SITS analysis approach can be modified to alleviate bias due to these confounders. The first approach relies on an unaffected dependent variable to account for the confounding factors while the second approach employs unaffected units (i.e., areas and households) to control for the confounders. Each of these approaches are described in more detail below.

Figure 2a. Short interrupted time series with linear trend and an unaffected outcome to estimate year-specific impacts



**Removal of bias using an unaffected or untreated outcome.** This approach entails using a measure that is not specifically targeted by the program but influenced by the confounders in the same way as the targeted primary outcome measure. Specifically, this approach utilizes a measure that (1) is not targeted by the treatment, (2) follows a trend similar to the targeted outcome in the pre-intervention time period, and (3) is likely to be affected by the confounders in the same direction as the targeted outcomes. Such measures are called “non-equivalent dependent variables”, or shortly “untreated”, “untargeted” or “unaffected” outcomes (Shadish, Cook, and Campbell 2002; Coryn and Hobson 2011; Trochim and Donnelly 2007) . Figures 4a and 4b demonstrate how such a measure can be incorporated into the SITS analysis to control for confounders. Specifically, Figure 2a adds an unaffected outcome to the SITS framework, sample averages of which are represented by green triangles. Similar to the affected outcome, a baseline trend is constructed for the unaffected outcome (captured by the solid green line), which is then extrapolated into the post-intervention period to serve as its counterfactual (dotted green line).

Figure 2b. Short Interrupted Time Series with Linear Trend and an Unaffected Outcome to Estimate the Average Treatment Effect



In the absence of any confounders we expect the post-intervention observations of the unaffected outcome to be on the counterfactual line; therefore, any deviations from the counterfactual line are attributed to the confounding factors. That is, the term in Figure 4a represents the effect of the confounding factors on the unaffected outcome at the first post-intervention time point. In this stylized example, we assume the scales and the pre-intervention trends of the affected and unaffected outcomes are the same; therefore, the difference between the deviations of the affected and unaffected outcomes from their respective counterfactuals ( for the first post-intervention observation, for the second post-intervention observation, etc.) yields a treatment effect estimate that is adjusted for (free of) the influence of confounders. Figure 2b shows how the unaffected outcome can be employed in the estimation of the average treatment effect across all post-intervention observations accounting for potential confounders.

We propose using *maize yield* as an unaffected outcome, as it satisfies the three conditions specified above. That is, we believe that maize yield per hectare may be used as an unaffected outcome measure for either of the two primary targeted outcomes (uptake of on farm storage and smallholders’ income [maize revenue]) as it (1) is not specifically targeted by the AgResults Pilot, (2) is expected to follow a similar trend to the three outcomes (which we hope to verify with extant data collected and maintained by Tegemeo Institute and the University of Washington), and (3) should be affected by at least some (if not all) of the same confounders. We vetted these assumptions with pilot implementers during the pilot launch workshop. The general conclusion was that yield may not be affected immediately but it is feasible that farmers may re-invest their increased incomes from pilot to improve their productivity. Another reason why the program could have an impact on yield is if the distributors also use the opportunity to raise awareness and increase adoption of other inputs. While all of these impacts are feasible, we believe these affects will be muted in the initial years of the pilot and are unlikely to materialize by the time we conduct our endline.

**Removal of bias using a comparison group.** The second and preferred approach to addressing confounding effects is through the use of a comparison group. In the current context, a comparison group would consist of areas and/or smallholders that (1) are not targeted or affected by the treatment, (2) have similar characteristics to treated areas and smallholders, and (3) are likely to be influenced by the confounders in the same way as the treated areas and smallholders. The way the comparison group is incorporated in the SITS design (which then is also called a comparative short interrupted time series or C-SITS) is very similar to the untreated outcome analysis as described above. Specifically, when estimating separate treatment effects at each post-intervention time point, the pre-intervention observations of the outcome measure are used to create a counterfactual trend (the green triangles in Figures 2a and 2b now capture the average outcome for the comparison group), which is extrapolated to the post-intervention period to serve as the counterfactual for the comparison group. Deviations from this counterfactual at each post-intervention observation are attributed to the confounders and are used to adjust the deviations of the outcome measure of the treated areas and smallholders from their counterfactual.

As discussed above, we think that the identification of a sufficient number of appropriate comparison areas and households for the current evaluation may be infeasible due to potentially large differences between targeted and not-targeted areas and because such an effort would be costly. We have explored the feasibility of a more cost-effective approach to comparison group construction, which entails looking for potential comparison group members in existing household-level datasets. Note that while this approach entails using the existing panel data to identify the comparison group households, these households would need be administered the baseline and post-program surveys in the exact same manner as the households in the treatment group to ensure that the two groups have comparable data measures. Therefore, this strategy would require obtaining the identities and contact information of the households in existing panel datasets order to administer additional follow-up interviews to treatment and comparison group households in the sample. While such datasets do exist, they have not been made available to the Abt team and therefore we have determined this approach is infeasible.

## Model specifications for the estimation of the overall impact of the AgResults pilot

This section presents prototypical model specifications that implement the SITS design framework to estimate the effect of the pilot on the two outcomes of interest. We start with simple models that do not explicitly control for confounding factors and then present models that incorporate an unaffected outcome to correct for such factors. These models will be estimated separately for all outcomes of interest – the two primary outcomes (smallholder uptake of technology and smallholder income), food security, smallholder’s access to technology and farmers’ knowledge and attitude on storage practices.

For all these outcomes the SITS design requires multiple measures of baseline and one post-intervention data point. These data points will be collected through two surveys administered to the smallholders: the baseline survey that will include recall questions to obtain three pre-intervention data points and the endline survey that will yield the post-intervention data point. We are concerned that obtaining reliable retrospective data on income will be challenging. We believe that the primary channel through which the pilot is expected to influence income is through a reduction in sell-cheap/buy-expensive maize transactions patterns among smallholders and/or the price per unit the smallholders receive when they do sell their harvest. Therefore, as a proxy for income, we intend to use maize revenue as the outcome measure that we will analyse using the SITS framework.

For each outcome measure, we will estimate separate models with data from the Rift Valley and the Eastern regions to obtain specific impact estimates for each outcome measure in each region. The data for these analyses will be obtained from surveys conducted with households in the treated areas. We intend to create a panel dataset that includes as many as three pre-intervention values of the outcome measure and one to two post-intervention values . As discussed in the next section, we intend to represent all counties in the two areas (nine counties in the Rift Valley and five counties in the Eastern region).

**Levels of nesting in the data**. Within each county, villages will be sampled randomly, and within each village, households will be selected randomly for the survey administration. Therefore, in each region, the nesting structure in the data will be of the following form: individual observations (or time) nested within households nested within villages nested within counties. The simplest SITS model specification that reflects this cluster structure is given in Equation 1 below:



where:

= Outcome measure for household *h* in village *j* in county *k* at time *t*. As mentioned above, this model will be estimated separately for all targeted outcomes of interest: two primary outcomes (uptake of on farm storage and sale price of maize) as well as food security and knowledge and attitude on storage practices.

= The counter for observations, and time=*t*=1, 2, and 3 denote the three pre-intervention periods, while time=*t*= 4 denotes the post-intervention period..

= Indicator for the post-intervention observation (i.e., equals one if *t*=4 and zero otherwise).

= Indicator (i.e., fixed effect) for county *m (m=1,2,…,M).* It equals one if *k=m* and zero otherwise.

= The *n*-th characteristic of household *h* in village *j* in country *k* at time *t* (e.g., household size, age and gender of household head, land holdings, sources of income). Note that we allow these characteristics to be time variant, but we should be careful not to include those that could be endogenous (i.e., affected by the intervention).

= The *m*-th characteristic of village *j* in country *k* at time *t* (e.g., distance to the main road, access to electricity, distance to markets). Note that we allow these characteristics to be time variant, but we should include only exogenous attributes.

= Random effect for village *j* assumed to be distributed with a mean of zero and variance of . This variance term captures the outcome variation across villages within a given county.

= Random effect for household *h,* which is assumed to be distributed with a mean of zero and variance of . This variance term captures the outcome variation across households within a given village.

= Residual associated with observations at time *t*, which is assumed to be distributed with a mean of zero and variance of . This term captures the variation in the outcome measures of a household across time.

In this model, county effects are modelled as fixed because all counties will be represented in the survey sample, while the village and household effects are modelled as random to reflect the sampling variability introduced by the sampling carried out at these levels. In Equation 1, captures the linear time trend in the outcome measure (which is essentially based on the change in the outcome measure during the baseline period), and captures the treatment effect at the post-intervention data point. This model can further be modified to accommodate more complex situations and data patterns, including non-linear time trends (e.g., with the addition of the square of the time variable).

**Addition of an unaffected outcome to the analysis.** As mentioned above, treatment effect estimates yielded by the models in Equations 1 and 2 will account for the secular time trends in the outcome measure but may be biased due to confounding if other influences on the outcome *Y* change simultaneously with the arrival of the treatment. We now turn to how an unaffected outcome, represented by , can be used to remove this confounding (as mentioned above, we are currently considering using *maize yield per hectare* as the unaffected outcome measure). First, consider the estimation of the pooled impact estimate across all post-intervention time points and the following model specification for , which is parallel to the specification in Equation 1:



In Equation 2, all right-hand side variables are defined as in Equation 1, and the three error terms (and ) capture the random errors at the village and household levels and the time-specific residual, respectively. In this specification, represents the linear time trend in the untreated outcome measure, while captures the pooled deviation from this trend for the post-intervention time, point which is attributed to confounders and considered to represent the effect of the confounding factors. Note that this is based on the assumption that is not specifically targeted by the intervention; therefore, the deviation of this measure from its baseline trend is fully attributable to confounding factors. Further, assuming the effect of the confounders on the treated outcome is proportional to the effect on the untreated outcome where the ratio of the two effects is equal to the ratio of the time trends, the impact estimate for the treated outcome that is adjusted for the confounders is given by:



In Equation 3, is essentially the impact on the targeted outcome that is free of confounding factors that affect both the affected and unaffected outcomes in the same direction.

Another approach for using the unaffected outcome is creating an adjusted version of the affected outcome that is free of the time trend and the influence of the confounding factors that operate in the post-intervention period. This approach entails estimating the time trends for the untreated and treated outcome measures using only the pre-intervention observations via the following models:

(4a)

(4b)

Once the time trends for the treated and untreated outcomes are obtained via models 4a and 4b, the modified version of the treated outcome measure is created as follows:



This essentially creates a de-trended version of the treated outcome measure that is free of confounding factors that influence both the treated and untreated outcome measures. The adjusted outcome measure is then used in the following impact model:

(6)

Note that the model specification in Equation 7 does not include the time trend, which has already been netted out in the construction of the dependent variable. In this model, captures the treatment effect estimate that is adjusted for the secular time trend in the treated outcome measure. This estimate should be numerically equivalent to the adjusted impact estimate in Equation 3 while being more precise than . The difference between the precision levels of the two adjusted impact estimates is due to the time trend variable () used in the estimation of and being correlated with . Such undesirable correlation decreases the precision compared to that of , which does not suffer from such a precision loss because the model in Equation 6 does not include the trend variable.

## Differential impact of the AgResults Pilot on subgroups of smallholders

In addition to reporting the overall average treatment effects, we propose to estimate the treatment effects for various subgroups of smallholders. We know that DFID is interested in exploring whether the impact of the pilot is different for female-headed households. Other potential subgroups of interest include smallholders who at baseline are (1) less credit-constrained, (2) have a higher level of education, (3) have a larger household (pool of potential laborers), or (4) have more (or more advanced) farming inputs. All of these groups may experience different intervention impacts than smallholders not in these categories. In order to investigate whether the pilot’s impacts on any of these subgroups are different than the rest of the sample, we intend to (1) estimate the impact models specified in the previous section separately for each subgroup of interest and the rest of the sample (i.e., estimating one model using only the households in the subgroup of interest and another model for the rest of the sample) and (2) compare the resulting subgroup-specific impact estimate to that for the rest of the sample.

As described in the next section, the survey sample has not been built at a scale to provide for confident analysis of subgroup-specific effects given that only a portion of the data can be used for each examined subgroup. But it will be feasible to detect impacts on subgroups of a large magnitude for outcomes.

# Statistical Power Analysis and Sampling Frame

This study relies on household-level data collected from a panel of small farmer households in each of the 14 counties where the pilot is active. Baseline data collection took place during July and August 2014, with a second round of baseline data to be collected in 2015 (owing to the fact that the program started later than projected and there is now an additional pre-program year on which we must collect baseline data), and endline data will be collected in 2019.

## Power Analysis

We conducted a series of statistical power analyses to estimate the number of villages and households that would be needed in the evaluation sample to be able to detect the targeted effect magnitudes with reasonable statistical power under the analysis approaches described above. The sample size requirement is determined based on the first outcome – uptake of storage, which is the most proximal outcome that the program aims to influence. We then present the effect size that can be detected for the second outcome (sale price of maize) with this sample size. Given our plans of conducting separate analyses for the two regions and the differences between the expected effects in the two regions (which is described in more detail below), we conducted the power calculations separately for the Rift Valley and the Eastern regions.

**Minimum detectable effect size (MDES) formula for the preferred impact analysis model.** The starting point for conducting a power analysis is to stipulate the underlying model specification that will be used to estimate intervention effects. In this case, given current uncertainties associated with securing a reasonable untreated outcome and/or comparison areas and household, we decided to be conservative and base the power analysis on the SITS model specification in Equation 1, which uses pre-intervention data points to form a linear baseline trend extrapolated into the post-intervention period to serve as the counterfactual, yielding a pooled treatment effect estimate across all post-treatment data points. Specifically, the following formula is used in the power analysis:

(7)

where:

captures the smallest impact estimate that can be detected with the given significance level, statistical power, and other parameters that effect the standard error of the impact estimate including the number of counties, villages, households, and pre- and post-intervention time points. In order to have a standard measure of power that can be used across all outcomes of interest, MDES is expressed in terms of the standard deviation of the outcome measure (i.e., it corresponds to an effect size).

= significance level (set to 0.05 for a 2-sided test).

= desired level of statistical power (set to 0.80 which implies an 80 percent probability of detecting a true effect).

= degrees of freedom which equals the total number of observations minus number of covariates and groups (counties, villages, and households).

= proportion of the outcome variance that lies across villages within counties.

= proportion of the outcome variance that lies across households within villages.

= proportion of the outcome variance that lies across multiple observations that belong to the same households.

= proportion of the outcome variance at the village-level explained by covariates included in the model.

= proportion of the outcome variance at the household-level explained by covariates included in the model.

= proportion of the within-individual outcome variance explained by covariates included in the model.

= square of the correlation between the *Post* indicator in Equation 1 (which yields the treatment effect estimate) and the time count. Note that this correlation is proportional to the sample size requirements (i.e., a larger correlation increases the number of household keeping the MDES constant).

= average number of villages per county.

= average number of households per village.

= number of observations per household.

= proportion of observations in the post-treatment period (which equals 0.25).

Equation 7 is based on the assumption that the outcome measure will be scaled so that its standard deviation is one and the corresponding impact estimate is expressed as an effect size. This also implies that sum of the all variance components (i.e., outcome variance that lies at the county, village, household, and time levels) is one:

(8)

Of these variance components, Equation 8 assumes that the outcome variance at the county-level () is fully explained by the county-level fixed effects included in the model represented in Equation 1.

**Required sample sizes.** We used the MDES expression in Equation 7 to calculate the required sample sizes of households in the Rift Valley and the Eastern regions. These analyses are based a number of parameter values and inputs:

* *Values for the village, household, and time-level variance components (, and ) and the R2 terms at these levels (, and ).* These values are obtained from secondary analysis of the panel data maintained by Tegemeo Institute and Michigan University. From this dataset, we analysed two measures to obtain the corresponding parameter values for our two outcomes: (1) number of 90 kg bags used for storage (for the uptake of storage solutions) and (2) revenue obtained from maize sold (as a proxy for income).
* *Anticipated program penetration rates of 6 and 18 percentage points in the Eastern regions and the Rift Valley region, respectively, obtained from the AgResults Pilot Business Plan (Dahlberg 2012).* Given that each “treated” household is expected to have access to four 90 kg bags and the standard deviation of the number of bags in the Tegemeo Institute dataset is close to four, the target penetration rates correspond to MDES estimates of 0.06 and 0.18 for the first outcome uptake of storage solutions (as measured by the number of storage bags each household has) in the Eastern region and the Rift Valley region, respectively.
* *Availability of three pre-intervention data points and one post-intervention data point for each outcome measure in each region.*
* *Maximizing the number of villages sampled.* This is desired since it would minimize the first component of the MDES formula in Equation 8 (village-level variance term scaled by the number of villages) and consequently decrease the MDES estimate. We believe that the optimum number of households that can be surveyed in each sampled village is approximately six because a smaller threshold may lead to the inefficient use of visiting time to villages, while a larger number would not maximize the number of villages included.

Based on these parameter values and assuming 3 data points before intervention and 1 after intervention, in the Rift Valley we expect that we would need as many as 540 households (with the sampling of all 9 counties, 10 villages from each selected county, and the sampling of 6 households from each selected village) to reliably detect the expected effect size of 0.18 on uptake of on farm storage. Due to its lower penetration rate and hence smaller expected average impact, in the Eastern region we estimate that a sample of 4140 households (5 counties, 138 villages per county, and 6 households per village) will be needed to detect the anticipated effect size of 0.06 in that region. These sample sizes would also allow us to detect similar effects on the second outcome of interest (income or sale price of maize): 0.16 in the Rift Valley region and 0.06 in the Eastern region.

As mentioned above, these power analyses are conservative in the sense that they do not incorporate the potential use of an untreated outcome or a comparison group to strengthen the analysis. It is also important to note that they do not apply to analyses of impacts on subgroups of interest such as female-headed households. If, say, one-quarter of all smallholders fall into this category, we can be confident of detecting impacts twice the size of those stated above for such households. For example, in the Rift Valley region true impacts on uptake of on farm storage in excess of 0.36 (in effect size units) could be ruled out should insignificant test results emerge, as could true impacts in excess of 0.12 (in effect size units) in the Eastern region. Thus, even without significant findings (which are unlikely without much larger subgroup survey samples than can be afforded) something important will be learned about effects on subgroups of all sorts—not just the illustrative female-headed households discussed here.

## Sampling Frame

In each county of the two regions, survey areas (sub-counties) were purposively selected based on their likelihood of being targeted by implementers for marketing and sale of improved OFS technology based on (1) sub-county level maize production and yield data and (2) implementers’ own indications of which areas they planned to target as stated in their applications. As noted, we sampled 10 villages from each county in the Rift Valley region, and 138 villages in Eastern region. The villages were randomly selected from a list of villages compiled from the 2009 Kenya census.

In the baseline survey, at least six households were sampled from each village. (The sampling plan called for six households per village, though in a handful of cases more households were sampled.) Survey enumerators used a random walk to select households, starting from a geographic landmark such as a main road junction or a permanent structure such as a church or bridge, and followed a pre-determined skipping interval. In most cases, this interval was five, but in some villages where households were particularly far apart, a lower skipping interval was used. Survey enumerators then screened households based on two criteria to ensure that the impact evaluation targets the same population as the implementers so as to accurately capture the true change in outcomes for the targeted population:

* The target household must have grown at least one crop in 2013 (the year for which data were being collected) that could be stored in an improved OFS container. The qualifying crops were barley, beans, cowpeas, green grams, maize, millet, njahi (a type of black bean), pigeon peas, and sorghum.
* The target household must have cultivated no more than 20 acres (in Eastern counties) or 35 acres (in Rift Valley counties). The purpose of this cut-off was to exclude “large” farmers from the sample, as AgResults targets smallholders[[2]](#footnote-2).

# Key Variables and Data Sources

The following table presents each final outcome variable along with its definition, the analytic method used to estimate changes (either SITS or before-after comparison) and the data source, including the relevant question number(s) from the household-level baseline survey.

|  |  |  |  |
| --- | --- | --- | --- |
| **Quantitative Sub-question (represented as Y in the model in Section 1.2 above)** | **Definition of outcome variable** | **Analytic method** | **Data source and baseline survey question number(s)** |
| **Evaluation Question 2** | | | |
| What is AgResults’ impact on smallholder uptake of on-farm storage solutions? | Binary variable indicating whether household has adopted each OFS technology by the time period in question | Time series analysis of ownership of OFS | M04\_A to M04\_J (household had ever used the specified OFS technology by the time period in question) |
| **Evaluation Question 3** | | | |
| What has been AgResults’ impact on smallholders’ income? | Continuous variable constructed as total maize revenue (price\*quantity)  (Note: maize revenue is used here as a proxy for income, as it is a more relevant gauge of income changes likely to be influenced by the pilot) | Time series analysis of maize revenue  (Note: maize revenue is considered to be zero if the household did not sell maize during that season) | I10 (quantity sold) \* I11 (price) |
| **Additional outcomes** | | | |
| What is AgResults’ impact on smallholder awareness of on-farm storage solutions? | Binary variable indicating whether respondent has ever heard of each improved OFS technology by the time period in question | Time series analysis of awareness of OFS | 2013: M01\_A to M01\_J (respondent is aware of the specified OFS technology for grain/pulse storage) |
| Change in sale price received for 9 crops that can be stored in on-farm storage solutions | Sale price for largest sale (KSH/kg) of maize and other grain/pulse crops | Time series analysis of reported sale price for maize; before-after comparison of reported sale price for non-maize crops  (Note: sale price is coded as missing if the household did not sell the crop) | 2013: I10-I11, I22-I24 (unit and sale price per unit for maize and 2nd most important grain/pulse crop)  2012: J02 (unit), J04 (price per unit) (maize only)  2011: J16 (unit), J18 (price per unit) (maize only) |
| Change in quantity of maize sold right after harvest | Quantity of maize sold (kg) after harvest | Time series of maize quantity sold right after harvest  (Note: quantity sold is considered to be 0 if the household did not sell that season) | LR 2013: I10  LR 2012: J02  LR 2011: J09 |
| Change in month of sale for 9 crops that can be stored in on-farm storage solutions | Month of sale | Time series analysis for maize; before-after comparison for non-maize crops | 2013: I08, I20 (month/year of sale for maize and 2nd most important storable crop)  2012: J03 (month/year of sale for maize only)  2011: J17 (month/year of sale for maize only) |
| Change in length of time between maize harvest and running out of maize stored for consumption (food security) | Number of months between end of harvest and running out of maize stored for consumption from that harvest | Before-after comparison of length of maize storage for consumption for adopters vs non-adopters | 2013: N03  2012: J06  2011: J20 |
| Change in quantity of stored grain / pulse that is lost during storage due to all causes combined | Total quantity of loss as estimated by respondent during total length of storage (kg) | Before-after comparison of reported losses for adopters vs non-adopters | LR 2013: L10 (quantity and unit of all losses) |
| Change in quantity of maize purchased for consumption | Quantity of maize purchased for consumption over the past 12 months (kg) | Before-after comparison of quantity of maize purchased for consumption for adopters vs non-adopters | 2013: N05-N06 (quantity and unit of maize purchased for consumption over past year) |
| Change in amount spent on maize for consumption | Amount spent on maize for household consumption over the past 12 months (KSH) | Before-after comparison amount spent on maize purchased for consumption | 2013: N05-N06 (quantity and unit of maize purchased for consumption over past year) |

Finally, we present here a list of covariates used in our SITS model (collectively represented as *w* in the model in Section 2.1 above):

|  |  |
| --- | --- |
| **Variable** | **Definition, units, and notes** |
| Region | Eastern or Rift Valley |
| County | County in which household resides |
| Household head gender | Binary indicating male or female household head |
| Household head age | Age in years |
| Household head highest education level | Highest education level completed |
| Highest education level in the household | Highest education level completed |
| Household head literacy | Binary indicating whether household head is literate |
| Household size | No. of HH members |
| Household had any members who live in another city to work | Binary indicating whether any member of the household lives and works in another location |
| Household is very poor | Estimated; see USAID Poverty Assessment Tool documentation for more information at http://www.povertytools.org/countries/Kenya/Kenya.html |
| Total land size (owned) | Household’s total owned land (ha) |
| HH rents any land (for any use) | Binary indicating whether household rents in any land (ha) |
| Total land size (rented in) | Household’s total rented-in land (ha) |
| Distance to nearest motorable road | Km |
| Distance to nearest tarmac road | Km |
| Distance to nearest main market | Km |
| Distance to nearest ag extension office | Km |
| Distance to nearest matatu/bus stop | Km |
| Total spent on non-HH labor | KSH |
| Total spent on planting materials/other farming inputs, long rains 2013 | KSH |
| Whether irrigated any plot, long rains 2013 | Binary indicating whether household irrigated any plot during the most recent growing season |
| Number of household person-days spent on farm labor, long rains 2013 | Number of days |
| Household needed any loans within the past year | Binary indicating whether household reported needing any loans during the past 12 months |
| Total off-farm income | Total household labor income (KSH) |

References

Bloom, Howard S. 2003. "Using "Short" Interrupted Time-Series Analysis to Measure the Impacts of Whole-School Reforms." *Evaluation Review* 27:3-49.

**Coryn, C. L. S.**, & Hobson, K. A. (2011). Using nonequivalent dependent variables to reduce internal validity threats in quasi-experiments: Rationale, history, and examples from practice. In S. Mathison (Ed.), Really new directions in evaluation: Young evaluators’ perspectives. New Directions for Evaluation, No. 131, 31-39. San Francisco, CA: Jossey-Bass.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008) Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, 27(4), 724–750.

Trochim, W. M. K, & Donnelly, JP (2007). *Research methods knowledge base*. Cincinnati, OH: Atomic Dog Publishing.

1. See, for example:

   Shadish, William R., Cook, Thomas D., Campbell, Donald T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston, MA, US: Houghton, Mifflin and Company

   Bloom, H. (2003). Using “short” interrupted time-series analysis to measure the impacts of whole-school reforms: With applications to a study of accelerated schools. *Evaluation Review, 27*(1), 3–49. [↑](#footnote-ref-1)
2. There was disagreement among various experts and officials at the Kenyan Ministry of Agriculture as to what amount of land holdings qualifies small versus medium or larger farms. Furthermore, at the time of the survey the AgResults pilot team had not yet determined the cut-off to be considered a smallholder. The Abt team’s earlier qualitative research also indicated that slightly larger and/or wealthier farmers may be more able and willing to buy improved OFS than truly “small” farmers. As such, the cut-off was set such that all small- to medium-scale farms would be included in the sample. A higher cut-off was set in the Rift Valley because farms there tend to be larger on average. [↑](#footnote-ref-2)