International Initiative for Impact Evaluation
Improving lives through impact evaluation

Pre-analysis plan (PAP)

The effect of demonstration plots and the warehouse receipt system on ISFM adoption, yield and income of smallholder farmers: a study from Malawi's Anchor Farms	TW4.1018
Authors	Annemie Maertens and Hope Michelson

1. Methods overview

1.1. Data sources

The data collection consists of: (i) a qualitative component, (ii) agronomic data collection, (iii) household questionnaires, and (iv) village questionnaires. In addition, we collect CDI monitoring data. The baseline and endline (in Fall 2014 and Fall 2018 respectively) include (i), (ii), (iii), and (iv); and the midline (Fall 2015) includes a shorter version of (iii) and (ii). At midline, we propose to collect (ii) and (iii). Table 1 gives an overview on (ii) and (iii).

Table 1: Overview of data collected (modules)

	Baseline	**Midline**
Agronomic data collection	Soil characteristics, plot history, GPS data of demonstration plots and selected farmers' plots Rainfall near demonstration plots Yields on demonstration plots	Rainfall near demonstration plots Yields on demonstration plots
Household questionnaires	Household composition and identification Landholding Assets Marketing and agricultural input/output for 2013-14 Time preferences Beliefs regarding ISFM Social networks Recall on ISFM adoption	Changes in household composition and identification Social networks Participation in CDI activities (New) Beliefs regarding ISFM and prices (Updated) Adoption plans and constraints to adoption Risk attitudes (New) Credit and insurance constraints (New) Knowledge about ISFM techniques (New) Marketing and agricultural input/output for 2014-15 (Updated) Time preferences (Updated)

1.2. Identification strategy

(A) Intent-to-treat effect

Basic specifications

Denote the outcome variable of interest of individual i in village j as as y_{ij}. Assuming that treatment and control households are similar at baseline, we first compute the IIT (Intention-to-Treat) estimate of the **average treatment** effects as follows:

$$y_{ij} = \alpha + \beta_1 T_j + \mu_j + \varepsilon_{ij} \quad \text{(1A)}$$

Where $T_j \in \{0,1\}$ indicates whether the individual belongs to a treatment village or not[1], μ_j is a village level error term while ε_{ij} is an individual level error term. As correlation between observations can result in incorrect standard error estimates, we cluster standard errors at the village level to allow for arbitrary correlation between village members which affects y_{ij}.[2]

We can extend the basic specification in (1B):

$$y_{ij} = \alpha + \beta_1 T_j + \beta_2 D_j + \mu_j + \varepsilon_{ij} \quad \text{(1B)}$$

Where $D_j \in \{0,1\}$ indicates whether or not the individual lives in a village where a demonstration plot was established. Note that specification (1B) assumes that the demonstration plots were placed in villages which are comparable to villages were no demonstration plots were placed. The baseline report confirms that this is indeed the case (even though these results need to be taken with a grain of salt due to the limited number of villages which have demonstration plots).

To **increase the precision** of the β_1 estimate, we include baseline household characteristics in specification (1) by using (2):

$$y_{ij} = \alpha + \beta_1 T_j + \gamma \bar{X}_{ij} + \mu_j + \varepsilon_{ij} \quad \text{(2)}$$

Where $\bar{X}_{ij}$ denotes a vector of individual level characteristics at baseline.

To the extent possible, we explore **heterogeneity in treatment effects** as a function of baseline household characteristics to better understand the mechanisms driving our results and how likely they are to generalize to other contexts. This yields specification (3):

$$y_{ij} = \alpha + \beta_1 T_j + \gamma \bar{X}_{ij} + \beta_2 T_j \bar{X}_{ij} + \mu_j + \varepsilon_{ij} \quad \text{(3)}$$

[1] In treatment villages, lead farmers were invited to attend a course in December 2014, and all farmers in the village farmer clubs were invited to attend demonstration on nearby demonstration plot sites throughout the season.

[2] As the numbers of clusters in the midline is on the small side (100), we will compare this standard error specification with a village–fixed effect with robust standard error specification. For a discussion, see also Cameron and Miller (2013): A Practitioner’s Guide to Cluster-Robust Inference.

Baseline differences between treatment and control

As noted in the baseline report, while treatment and control villages are comparable in terms of observable village-level characteristics, households in treatment and control villages are not comparable across some of the observable dimensions, notably the gender of the household head and the number of household members.

This is likely due to the fact that in treatment villages, both club members and non-club members were sampled, specifically, we sampled 5 club members and 5 non-club members in each treatment village. As CDI club membership was self-selected, we could expect CDI club members to be different from non-club members, both along observable and non-observable characteristics. This implies that, unlike in the control villages where no self-selection into CDI clubs happened[3], the sample drawn in the treatment villages is unlikely to be representative of the village as a whole. Figure 1 explains this point. Denote by X, the unobservable along which households in the village sort themselves. Assume that if X is larger than a specified cut-off value, denoted $\bar{X}$, then the household participates in a CDI club, while if X is smaller than $\bar{X}$, the household does not participate in a CDI club. Now, assume that the distribution of X follows a normal distribution with mean 0.

Indeed, only if $\bar{X} = 0$, the sample selected in the treatment village is a representative sample of the village as a whole.

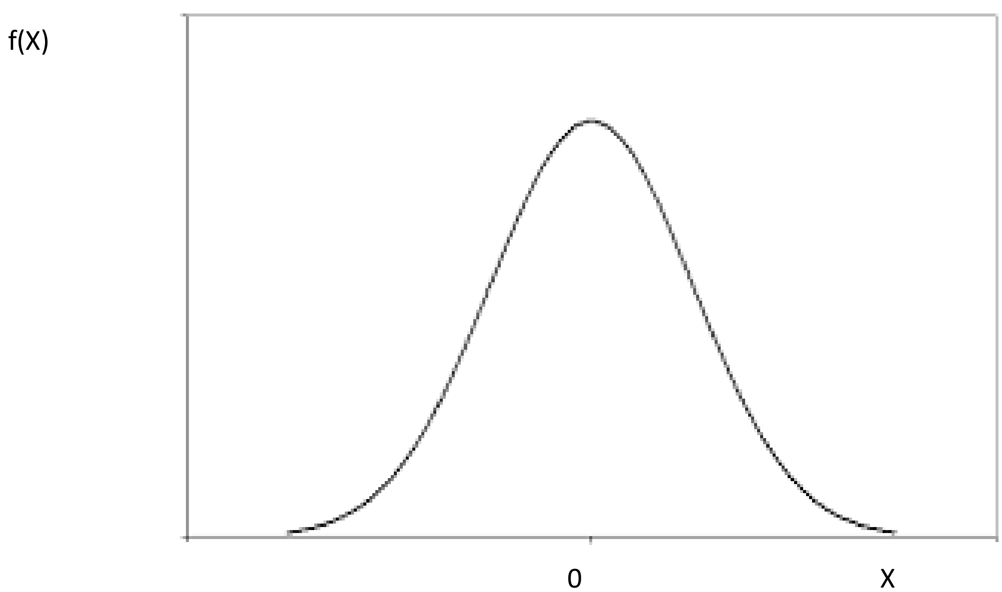

Figure 1: Self-selection into CDI club membership

[3] Unfortunately, logistical limitations did not allow CDI to form clubs in the control villages. This is because the formation of clubs coincided with a first information session about the CDI program, and hence the treatment and club-formation are intrinsically linked to each other. If this were not the case, one could simply compare club members in treatment villages to club members in control villages; as well as non-club members in treatment villages and non-club members in control villages.

Hence, we propose to proceed as follows. The method outlined next is inspired by the method of Propensity Score Matching. For the treatment villages, we use the baseline observable characteristics to predict self-selection into a CDI through a Probit specification [4] as in specification (4):

$$P\left(club_{ij,treatment} = 1\right) = \alpha + \beta_1 \bar{X}_{ij} + \mu_j + \varepsilon_{ij} \quad (4)$$

Using (4), we predict CDI club membership in the control villages as outlined in specification (5), where $\hat{\alpha}$ and $\hat{\beta}_1$ of (5) are the estimates obtained from regression specification (4). Note that the fact that standard errors in (4) were clustered at the village-level does no longer enter specification (5) as the clustering of errors only affects the precision of the β_1 estimate and not the β_1 estimate itself.

$$P\left(club_{ij,control} = 1\right) = \hat{\alpha} + \widehat{\beta_1} \bar{X}_{ij} \quad (5)$$

We then proceed to define a cut-off rule. For instance, if a farmer's probability is larger than 0.5, we would assume club-membership, while if it is smaller than 0.5 we would assume non-membership. Using this cut-off rule, we can divide the farmers in the control sample into two groups: the group of the "would-be" club members, and a complementary group of not "would-be" club members.

We can then proceed and estimate the effect of the treatment for CDI club members and CDI non-club members through two separate specifications as in (6A) and (6B), i.e., we effectively run these two regressions for the two mutually exclusive sub-samples constructed: (i) the sub-sample of club-members in the treatment villages and predicted club members in the control villages; and (ii) the sub-sample of non-club members and predicted non-club members in the control villages:

$$y_{ij,club} = \alpha + \beta_{1,\text{club}} T_j + \mu_j + \varepsilon_{ij} \quad (6A)$$

$$y_{ij,non,club} = \alpha + \beta_{1,\text{non-club}} T_j + \mu_j + \varepsilon_{ij} \quad (6B)$$

[4] Alternatively, a linear probability model can be used.

As an alternative to this Propensity-Score-Matching inspired approach, a one can use a traditional Ordinary Least Squares (OLS) regression approach to control for differences in baseline characteristics as in specification (2) or a Household Fixed Effects Approach to control for both observable and unobservable household-level time-fixed effects.

However, for this approach to result in an estimate of the average program effect, one would need to assume that the effect size of the program for club members and non-club members is comparable. As club-members can be expected to join the club in the first place because they exactly expect a higher beneficial effect of the program compared to non-club members, this condition is unlikely to hold true. Footnote [5] further details this somewhat technical issue. Note also that adding club-membership status as an interaction term with the treatment variable is unlikely to solve this issue, as club-membership is likely to be correlated with the error term.

(B) Average Treatment Effect on the Treated

Basic specification

The previous specifications ignore self-selection in terms of attending the trainings when one is invited. This decision is likely to be driven by unobservable individual characteristics. Hence, we estimate the ATT (Average Treatment Effect on the Treated), or in our case, the average effect on the compliers (assuming that no-one in the respective control groups are able to take up the treatment).

To estimate ATT, we will use a **two-stage least squares instrumental variable strategy**, where the randomized offer of treatment will serve as an instrument for the treatment. Note that the first stage regression is of interest in and by itself to understand the constraints to participation in CDI activities (by interacting household-level baseline characteristics with the treatment in the first stage). This yields first-stage specification (7) and second-stage specification (8):

$$P\left(participate_{ij} = 1\right) = \alpha_2 + \beta_2 T_j + \mu_j + \varepsilon_{ij} (7)$$

---

[5] Imagine, for instance, that use a fixed effects approach and regress changes in the dependent variable on the treatment status, effectively differencing out both unobservable and observable time-invariant effects.

$$\Delta y_{ijt} = \alpha + \beta_1 T_j + \Delta\varepsilon_{ijt}$$

However, as we expect the effects of the program to be different for club-members as opposed to non-club members (one could imagine, for instance, that club-members exactly join the club for that reason, because they expect a large beneficial effect), the estimate of $\hat{\beta}_1$ would depend on the exact ratio of club and non-club members sampled in the treatment village. As this ratio was somewhat arbitrary set at 50%, there is no reason to expect that the estimate of $\hat{\beta}_1$ would be capturing the average treatment effect. However, if one knows the sampling weights, i.e., the number of (non) club members in the sample to the population, one could apply sample weights to (partially) correct for this bias (see the 2013 NBER study by Solon et al. 'What are we Weighting for?' http://www.nber.org/papers/w18859). These sample weights are known for a sub-set of the villages: namely for those were club-membership was perceived not to be restricted in which case the total number of club-members represents the population of club members.

$$y_{ij} = \alpha + \beta_1 \widehat{Participate}_j + \mu_j + \varepsilon_{ij} (8)$$

(Note a slight abuse of notation as we did not "rename" the error terms)

Take up of ISFM technologies

We note that the eventual welfare effects will depend on the links between ISFM technology adoption and income/welfare; i.e., if one adopts the promoted ISFM technologies, what are the expected effects on income and other measures of welfare? However, as we expect the effects of the program on technology adoption at midline to be limited, we postpone this welfare analysis to the endline stage. The endline analysis will hence contribute to a large literature in agricultural economics on the welfare effects of the adoption of new technologies. As not all farmers who visited the demonstration plot sites and participated in the training will adopt the wide arrange of technologies, we will again have to use an instrumental variable strategy to identify the effect of adoption on welfare, using the treatment assignment as the instrument.

Spillovers between treatment and control

As noted in the progress report, one can expect some spillovers between treatment and control villages. Using CDI program participation data during 14-15, household data collected at midline and GPS data, we intend to first document the extent of the spillovers through both CDI administrative data as well as midline data collected the households.[6]

If spillovers do appear significant, we intend to study them in detail, using specification (9), which is a variation of specification (1) which uses the sample of control villages only where S now indicates a measure of spillover.

$$y_{ij} = \alpha + \beta_1 S_j + \mu_j + \varepsilon_{ij} \quad (9)$$

To ensure a consistent estimate of β_1 in specification (9), we propose to use a GPS based distance measure to nearby treatment villages to instrument for our measure of spillover. Which measure of spillover we will use is still to be decided, but can range from "having heard about the CDI program" to "having visited a CDI demonstration plot".

1.3. Treatment assignment in 2014-15

As outlined in detail in the progress report, in treatment villages, lead farmers were invited to attend a course in December 2014, and all farmers in the village farmer clubs were invited to attend demonstration on nearby demonstration plot sites throughout the 2014-15 season.

[6] Current CDI participation data as discussed in the baseline report reveals that the extent of spillovers is much more limited as initially suspected. See the baseline report for details. We have yet to process the GPS data collected at baseline.

1.4. *Sample selection in 2015-16*

Recall that due to budget restrictions, the midline data collection will only revisit 100 out of 250 villages that were visited during baseline. The endline data collection, will revisit all of 250 villages. We propose to select an equal number of treatment and control villages for the purpose of the midline data collection, covering all 10 farmers in each village which were covered at baseline. We propose to include all demonstration plot villages in the sample, as well as all villages that were covered by the soil science team at baseline – this is a total of 50 villages which are preselected. The remaining 50 villages will be drawn – at random – from the remaining set of treatment and control villages.

2. Main hypotheses under investigation

Recall that the overall goal of the project is to evaluate **the effect of AFM interventions on farmers' agronomic income**, but given lack of power, we focus testing the theory of change which relates to secondary outcomes: adoption of ISFM practices (crop rotation, fertilizer, optimal practices), yield and acreage of soy and maize, dietary diversity, beliefs about the costs/benefits of ISFM technologies, soil fertility, and prices and time of sales.

Using the midline data, we can investigate our first set of hypothesis as outlined in the theory of change section in the project proposal.

1. Exposure to demonstration of ISFM sowing and harvesting practices (sowing density, incorporation of residues, use of seed inoculant) *increases adoption of these practices, improves soil properties and yields*

2. Exposure to demonstration of optimal use of fertilizer *increases fertilizer use and increases yields*

3. Exposure to demonstration of soy-maize rotation *decreases total maize production and affects maize consumption and diet diversity*

4. Exposure to demonstration of soy-maize rotation *increases soy adoption, increases rotation thereby improving soil properties, increases farmers' exposure to soy market price fluctuations*

In addition, we investigate the overall effect of the treatment effect on agricultural income in 14-15.

3. Hypotheses, hypothesis testing, and heterogeneities of interest

Below, we provide details on the hypotheses to be tested. All these hypotheses will be tested using 2-sided tests using the framework outlined in Section (1) of this Plan. Standard errors will be clustered at the village level.

3.1. *Hypotheses*

1. Exposure to demonstration of ISFM sowing and harvesting practices (sowing density, incorporation of residues, use of seed inoculant) *increases adoption of these practices, improves soil properties and yields*
   Intermediate indicator: adoption of ISFM practices in 14-15, planned adoption of ISFM practices in 15-16 and perceived constraints to adoption, subjective beliefs on per acre yields of maize, groundnut and soy as well as knowledge about ISFM technologies beliefs/knowledge in 2015, yields of maize, groundnut and soy during 14-15.
   Modules: Input/output, landholding, soil tests, knowledge/beliefs
   Note: As the effects on soil properties are long term, we postpone looking at this indicator until the endline data.

2. Exposure to demonstration of optimal use of fertilizer *increases fertilizer use and increases yields*
   Intermediate indicator: use of mineral fertilizer in 14-15, subjective beliefs on per acre yields of maize, groundnut and soy as well as knowledge about ISFM technologies beliefs/knowledge in 2015, yields of maize, groundnut and soy during 14-15.
   Module: Input/output modules

3. Exposure to demonstration of soy-maize rotation *decreases total maize production and affects maize consumption and diet diversity*
   Intermediate indicators: maize acreage in 14-15
   Modules: Input/output, consumption
   Note: As the effects on consumption and diet diversity are longer term, we postpone looking at this indicator until the endline data.

4. Exposure to demonstration of soy-maize rotation *increases soy adoption, increases rotation thereby improving soil properties, increases farmers' exposure to soy market price fluctuations*
   Intermediate indicator: soy acreage in 14-15, subjective beliefs on per acre yields of maize, groundnut and soy as well as knowledge about ISFM technologies beliefs/knowledge in 2015, market prices in 14015
   Module: Input/output, knowledge/beliefs, marketing and soil tests
   Note: As the effects on soil properties are long term, we postpone looking at this indicator until the endline data.

3.2. *Heterogeneous treatment effects and subgroup analyses*

A primary objective of the study is to **understand how starting variation in farmer and household characteristics and endowments is related to intervention uptake and outcomes**. Hence, we will study the role of baseline heterogeneity in credit access, land endowments, human capital (education of household head) and available labor (household size), asset wealth (asset index) as well as time preferences.

To the extent possible (as noted in the baseline report, there are only 18% female headed households in the sample), we explore gender as **female-headed households** can often face special labor and cash constraints compared with male-headed households.

In addition, we will measure two sources of critical variation generally unobserved (and therefore generally left out of analyses): **baseline soil fertility/structure and farmer beliefs about ISFM**. Because application of external inputs has been found to be considerably less effective on soils of poor quality and because farmers likely know or suspect the quality of their soils, it is critical to have measures of soil fertility to properly analyze adoption and the effects of adoption of new agricultural technologies. Moreover, because poor quality soils are often correlated with low-household wealth, failure to control for these often-unobserved variables can lead to faulty inference about why a household has failed to adopt a new agricultural technology. Similarly, heterogeneous starting farmer beliefs about expected costs and benefits of ISFM and commercial soy production are relevant to the study in two ways: first, because starting beliefs may critically influence farmer adoption and investment in new agricultural technologies. Second, because we will measure change in the beliefs themselves over time, comparing change between the control and treatment farmers to understand how exposure to the AFM affects farmer knowledge and expectations of costs and benefits.

Using the rainfall data we collected, and the longer-term series collected from the Department of Meteorology in Malawi, we might be able to shed light on the **relation between the effect size and weather**. Indeed, as discussed in detail in the progress report, a primary external threat to the realization of impact is the weather. A bad agricultural season could impact not only the effectiveness of the demonstration plots but, of course, farmers' own investment decisions, use of credit, and outputs and yields in both that agricultural season and the following season.

4. Addressing less-than-complete data

4.1. Addressing survey attrition

We expect survey attrition from base to midline to be rather limited, as only one year has passed. Nevertheless, we will check for non-random attrition and its potential effects using the framework outlined by Fitzgerald, Gottschalk and Moffitt (1998).[7]

4.2. Addressing item non-response

Similarly, while item non-response is very limited in the baseline (less than 0.1 percent for most variables), we propose to drop the observations which involve non-response or replace a missing variable by the sample average. Note that this method is only valid if no selection is suspected – which is something we intend to investigate if item non-response is substantial.

[7] Fitzgerald, J., P. Gottshalk and R. Moffitt (1998) 'An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study on Income Dynamics', Journal of Human Resources 33: 251-99

5. **Variables for controls, stratification and key sources of heterogeneity**
We laid out only the main outcome variables below, and not the intermediary indicators. Control variables mentioned above will also be used to look at heterogeneity, to the extent that the midline sample size will allow

Hypothesis group	**Indicate if *"treatment," "outcome, or "covariate"***	**Var name**	**Data source**	**Type of variable (binary, categorical, continuous)**	**Transformation of variable required**	**Plan for constructing an index, if needed**	**Plan for addressing outliers**	**Will missing data be imputed?**	**When will this variable be dropped?**
All	Treatment	Treatment	HH/B	Binary	NA	NA	NA	NA	Not dropped
(1)	Outcome	Adoption	HH/M	Binary/continuous	TBD	NA	TBD	NA	Not dropped
(2)	Outcome	Fertilizer	HH/M	Binary/continuous	TBD	NA	TBD	NA	Not dropped
(3)	Outcome	Maize acreage	HH/M	Continuous	TBD	NA	TBD	NA	Not dropped
	Outcome	Maize production	HH/M	Continuous	TBD	NA	TBD	NA	Not dropped
(4)	Outcome	Soy acreage	HH/M	Continuous	TBD	NA	TBD	NA	Not dropped
(4)	Outcome	Planned rotation	HH/M	Binary	TBD	NA	TBD	NA	Not dropped
(4)	Outcome	Soy market price	HH/M	Continuous	TBD	NA	TBD	NA	Not dropped
All	Control	Soil texture/structure	Soil	Continuous	TBD	NA	TBD	NA	Note: limited sample only
All	Control	Beliefs	HH/B	Continuous	No	NA	Drop 5%	NA	If midline reveals issues with baseline elicitation (see baseline report)
All	Control	Gender	HH/B	Binary	No	NA	No	NA	Not dropped
All	Control	Land owned	HH/B	Continuous	Log	NA	Drop 1%	NA	Land or asset index
All	Control	Education	HH/B	Continuous	No	NA	No	NA	Not dropped
All	Control	Number of adult HH members	HH/B	Continuous	No	NA	No	NA	Not dropped
All	Control	Assets	HH/B	Continuous	Yes	Yes, see baseline report	Drop 1%	NA	Land or asset index
All	Control	Time preferences	HH/B	Continuous	Discretize	Yes, see baseline report	Drop 1%	NA	If midline reveals issue with baseline elicitation (see baseline report)