Market Safe Pre-Analysis plan

November 8, 2017

Vivian Hoffmann, Sarah Kariuki, Janneke Pieters and Mark Treurniet

1. Introduction

Adoption of improved agricultural technologies in developing countries may be limited by common informational and market inefficiencies. In this study, we examine the role of such inefficiencies in the adoption of a new food safety technology, Aflasafe. Aflasafe has been shown to reduce aflatoxin contamination in maize by approximately 90%. However, this technology faces several barriers to widespread adoption. First, since Aflasafe is applied while the crop is still growing, its use increases farmers' exposure to yield risk. In the absence of a market for crop insurance, this may limit its adoption. Second, the lack of price incentives for food safety in markets served by smallholder maize producers in Kenya may constrain the adoption of food safety technologies by these farmers. In this study, we explore the impact on adoption of Aflasafe of 1) bundling the product with a rainfall index based money-back guarantee, and 2) access to an output market that rewards aflatoxin-safety.

The rest of this plan is organized as follows: Section 2 describes the study design; Section 3 describes the data sources used in the analysis; Section 4 lists the hypotheses to be tested; Section 5 describes the definition of variables to be used in the analysis; Section 6 details the estimation strategy; and Section 7 describes testing for heterogeneous effects. Finally, Section 8 describes how we will deal with spillovers and attrition.

2. Study design

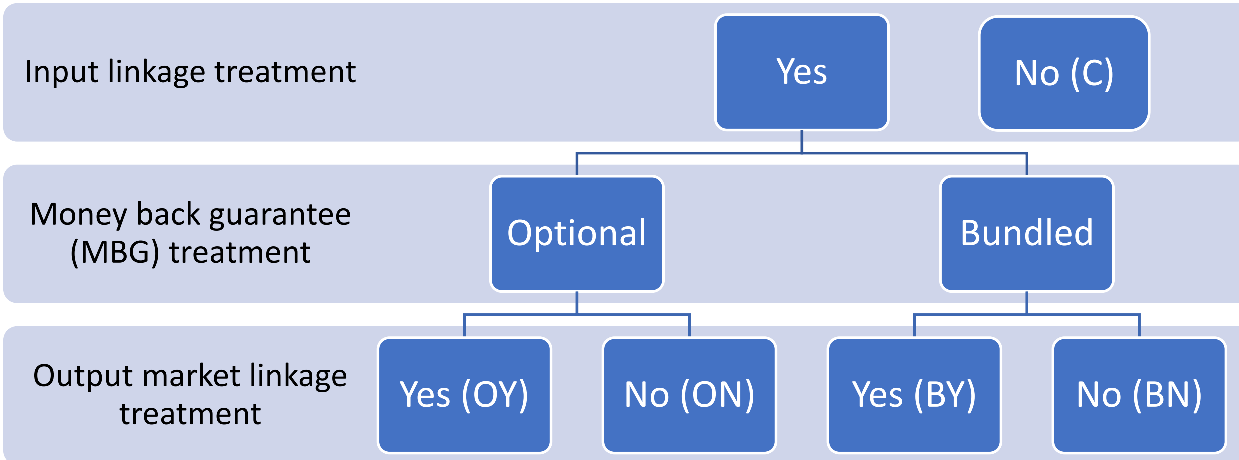
2.1. Experimental treatments

The study will contain 3 interventions: an *Input Linkage* treatment, a *Money Back Guarantee* treatment and an *Output Market Linkage* treatment. The research design is graphically presented in Figure 1.

The *Input Linkage* treatment groups will receive information on the benefits of Aflasafe and instruction on its use. They will also be given an opportunity to buy Aflasafe, which is not currently available in the study area.

These groups will further be divided into two treatment conditions concerning money back guarantee (MBG): an *Optional* condition, offered the option to buy Aflasafe either with or without the MBG, and a *Bundled* condition, in which Aflasafe is only offered in combination with the MBG. The *Money Back Guarantee* treatment will be crossed with the *Output Market Linkage* treatment, which consists of a linkage between the farmer group and a buyer that pays a premium for safe maize. Hence, in total there will be five categories, including the pure control group (C), and four treatment groups (OY, ON, BY, and BN).

Figure 1. A graphical presentation of the research design



2.2 Population

The population for this study consists of maize farmers who are members of existing farmer groups in Meru, Embu and Tharaka Nithi counties, Kenya. A list of 250 existing groups in the study area was acquired through the Cereal Growers' Association (CGA) and the Ministries of Agriculture in the three counties. From April to August 2017, all 250 groups were visited and lists of their members were obtained.

2.3 Treatment assignment

Farmer groups are drawn from the list of 250 farmer groups described above. Based on power calculations, we allocated 160 farmer groups across the four treatment groups and the control group as follows: 38 to OY, 38 to ON, 38 to BY, 38 to BN and 8 to C.

To limit spillovers across money back guarantee treatments, we created comparable, but geographically distinct, clusters of farmer groups within each of the study counties (4 in Meru, 2 in Tharaka-Nithi and 2 in Embu) and subsequently assigned these clusters to either the bundled or the optional MBG treatment. Subject to a minimum geographical distance of 5 km between clusters, we aimed to select similar farmer groups into the clusters within each county. Similarity was defined based on the Euclidean distance in the six-dimensional space formed by the standardized values of variables listed in section 5.4.

To this end, we first dropped farmer groups close to the county borders to achieve a minimum distance of groups in different counties of at least 5 km. Subsequently, we excluded any groups within a 5 km bands dividing the remaining groups into similarly-sized clusters. The direction of this band was selected to minimize the Euclidean distance between the groups on either side of it. From the remaining groups, we then iteratively selected matched pairs across each cluster with the lowest Euclidean distance into the sample. To ensure that the MBG treatments were spread out geographically, we manually decided which clusters would receive the same treatment. Finally, we randomly assigned the bundled MBG to one of the two groups of clusters.

Within each of the money back guarantee clusters, the market linkage treatment was randomly assigned at the village level (62 villages to Market Linkage and 62 villages to No Market Linkage).

Pure control groups were selected as the 8 nearest geographical neighbors to any BY group, stratified by county approximately in proportion to the total number of groups on the initial list.

3. Data sources

Farmer group census: A short survey of all 250 farmer groups on the initial list was conducted during meetings with these groups in April-August 2017, for the purposes of sample selection, treatment assignment and balance checks. Data on the groups' geographical location, and members' familiarity with weather insurance, awareness of aflatoxin, use of agricultural inputs, and maize production and marketing were collected. Lists of the groups' member were also obtained.

Baseline survey: Baseline survey data both at the individual farmer level and at the farmer group level was collected in September-October 2017, the beginning of the maize season in the study area. Six farmers per group were randomly selected from the lists obtained during the group census to participate in the individual level survey. In case any of the selected farmers were not available, replacements were selected from an ordered list of six additional farmers. The group leader was interviewed using the group level questionnaire.

Aflasafe application and sporulation: A subset of farmers who have purchased Aflasafe will be visited by IITA to observe how the product has been applied and whether it has sporulated (a requirement for it to be effective).

Follow-up survey: A follow-up survey with the same respondents will be conducted after the harvest season, around March-April 2018. Samples of maize stored for household consumption and later sale will also be purchased from these farmers, and will be tested for aflatoxin and/or Aflasafe fungal strains.

Administrative data: Additionally, data on training attendance and Aflasafe purchases will be collected during sales of Aflasafe in November and early December 2017. Data collected will include the number of farmers present at the training. For all farmers who purchase Aflasafe, the following will be recorded: name, gender, land area under maize, and the number of packets of Aflasafe (with and without the MBG) purchased. For the farmer groups receiving the output market linkage treatment, data on the volumes aggregated and price received shall be collected, starting at the time of aggregation in February-March 2018. Data on insurance activations and payouts will be collected by Acre Africa, from the time of Aflasafe application through the end of the season (November 2017-March 2018).

4. Hypotheses to be tested

Hypothesis 1: Providing farmer groups with information on aflatoxin and access to Aflasafe reduces aflatoxin contamination in farmers' maize stored for household consumption.

Hypothesis 2: Adoption of Aflasafe is higher when farmers are linked to a market that rewards aflatoxin-safe maize.

*Hypothesis 3***:** Adoption of Aflasafe is higher when farmer groups are offered a money back guarantee bundled with the Aflasafe, compared to an optional money back guarantee.

A rational decisionmaker's Aflasafe adoption would be weakly higher under the optional money back guarantee treatment. However, from previous studies we know that although farmers invest more when insured, willingness to pay for insurance is low. We therefore expect that many farmers will buy Aflasafe without the money back guarantee when given the option. Since farmers invest more when they are insured, we hypothesize that the adoption of Aflasafe will be higher under the bundled money back guarantee treatment.

Hypothesis 4: Investment in food safety is higher for own consumed food than for sold food in the absence of market rewards for food safety, but this gap shrinks or could even reverse in the presence of market incentives. To test this hypothesis, we will compare either or both: 1) the difference in aflatoxin levels in maize for sale, minus aflatoxin levels in maize for home consumption in the non-market linkage groups versus the market linkage groups; 2) the difference in Aflasafe usage between maize stored for home consumption and maize stored for sale among farmers in the market linkage groups versus the non-market linkage groups.

5. Variable definitions

5.1 Outcome variables

1. Adoption of Aflasafe

Farmers' adoption decisions will be measured by a dummy variable, equal to 1 if the farmer adopted and 0 if the farmer did not adopt. The intensity of adoption will be measured as the amount of Aflasafe purchased. Both the binary and intensity adoption variables will be constructed using Aflasafe sales data.

2. Aflatoxin contamination and Aflasafe usage:

Samples of maize will be taken from the following three sources:
1) Maize stored for household consumption
2) Maize stored for later sale
3) Maize aggregated for testing and sale through the study

Samples will be tested for aflatoxin using a quantitative test with an upper detection limit of 150 ppb. In the event that more than 5% of samples are at or above the upper limit of the detection range, these will be diluted and re-analyzed, to a maximum detection level of 400. A microbiological test will be used to assess whether Aflasafe was used on a batch of maize. Selection of one or both aflatoxin contamination and/or Aflasafe usage depends on the relative costs and efficiency of these two indicators for assessing farmer behavior.

5.2 Individual level characteristics measured at baseline, to be used as controls

<u>Characteristics measured through the individual baseline survey:</u>

1. Age of the farmer (completed years) (B6, row 01)

2. Sex of the farmer (0=female, 1 male) (B5, row 01)
3. Years of education completed by head (B7, row 02)
4. Relationship with the head (1=household head; 0=otherwise)
5. Asset index (sum of dummy variables indicating ownership of each asset listed in sections C1 and C2.
6. Total land under maize previous season, Winsorized (D1)
7. Maize harvest (kg) previous season, Winsorized (G2)
8. Maize marketing: whether sold any maize last season (J1)
9. Total expenditures on agricultural inputs (E4), including labor, previous season plus the total labor expenditures on maize calculated from E8-E10, Winsorized.
10. Propensity for social learning (binary variable indicating whether the number of people the farmer discusses technologies with in the village is below or above the sample median).
11. Aflatoxin knowledge at baseline: an index constructed as the 0.6 times the z-score of the sum of dummy variables indicating correct or affirmative answers to questions K1, K2, and K3, plus 0.2 times the z-score of the number of correct responses to K4 plus 0.2 times the z-score of the number of correct responses to K5, where correct or affirmative answers are defined as follows: K1=1, K2=1, K3=any of 1,2,3,4, or any correct 'other' response, K4=any pre-coded response aside from 10, or any correct 'other' response, K5=any pre-coded response aside from 7, or any valid 'other' response.
12. Risk aversion: Information to assess risk aversion was collected using survey responses during the individual level baseline survey. We adapted a module developed by Falk et al. (2016) to assess risk preferences using survey data. This involves a qualitative/subjective assessment by the respondents on how willing (unwilling) they are to take risks (P1e and P1f), as well as a quantitative assessment through a set of hypothetical questions involving 3 interdependent choices between a lottery and a sure amount (P2). A measure of an individuals' risk preference will be indicated by the row in which the respondent switches from preferring the lottery to preferring the sure payment. The higher the switching row, the more risk loving (less risk averse) a respondent is. The risk aversion indicator (qualitative or quantitative) that explains the most variation in the Aflasafe adoption decision will be included as a control.

12\. Knowledge / experience with insurance (the sum of indicators for whether the respondent has ever heard of money-back guarantee; has heard of crop, livestock or weather insurance, and has ever subscribed to weather related insurance.

13\. Trust: Trust will be assessed using survey responses during the baseline survey with the individual respondents. We will define "trust" as the mean of the following two variables: 1) the mean, multiplied by 2/3, of the following two variables: a) standardized proportion of group members trusted with unobservable task (N7/N6), b) standardized indicator for whether the group leadership is trusted (N8); and 2) the standardized mean of responses to questions P1B, P1C, and P1D.

Characteristics measured through the group census:

1. Whether the individual was present at the group meeting at which the census took place.
2. Farmer sex (coded based on name)

5.3 Group Level Characteristics to be used as controls

Characteristics measured through the baseline group survey:

1. County indicator

2. Trust: The mean, at the group level, of the individual trust index described above.

3. Group capacity: Group capacity will be measured as the standardized mean of the following 5 variables: 1) standardized size of the group (Group survey B1a), 2) standardized number of members who came to the initial meeting (Group survey B25), 3) Standardized value of a dummy indicating that the group does or has conducted any joint marketing activities (input purchases or output aggregation), constructed as the maximum of the following variables: Group survey B8=2, Group survey B8=3, Group survey B10, Census b3=1, Census b4=1, 4) standardized value of a dummy for members own initiative (Group survey B5), 5) standardized value of a sub-index of the (categorical) proportion of members participating in group's main activities (B9) and meetings (B17), constructed by first standardizing each of these variables and then taking their mean.

4. Gender: proportion of total members who are female (Group survey B1b / B1a)

Other group-level variables

1. Rainfall indicators (insurance triggers provided by ACRE), which will be matched to respondents based on the GPS coordinates of the group meeting location.

5.4 Variables used for balancing of money-back guarantee treatments

Variables used during cluster formation (all measured through the group census, all standardized):

1. Group size (B35)
2. Mean amount fertilizer used (B34e)
3. Proportion of members using fertilizer (B6/B5b)
4. Mean harvest normal season (B34c)
5. Sum of proportions of members that know crop or weather insurance (B8/B7b), have ever been trained on crop or weather insurance (B10/B9b), have ever purchased or been given crop or weather insurance (B12/B11b) and know of anyone who received a payout from crop insurance (B14/B13b)

6. Sum of proportions of members that have ever heard of aflatoxin (B24/B23b) and members that normally buy maize from any source and would pay extra for maize that they knew was aflatoxin-safe (B30/B29b)

We checked the variables used in cluster formation balance across treatment groups. In addition, we also checked the following variables for balance:

1. Dummy indicating whether the group had ever aggregated maize for sale before (B3)
2. Mean proportion of bags of maize sold from Feb/March harvest in a normal season (B34d/B34c)
3. Rainfall indicators (insurance triggers provided by ACRE)

6. Estimation strategy

6.1 Aflatoxin contamination

Hypotheses 1

To assess the effect of the input treatment on aflatoxin contamination of maize stored for household consumption, we will use the following equation, limiting the sample to farmers in groups assigned to the control group and the nearest-neighbor match of groups in the treatment group.

$$Aflatoxin_{ijt} = \alpha_1 + \beta_1 \cdot Treated_j + \gamma P_j + \varepsilon_{ij1} + \varepsilon_{j1}, \qquad (6.1)$$

Where $Treated_j$ is a dummy equal to 1 if group j receives the input linkage treatment, $Aflatoxin_{ijt}$ is the level of aflatoxin measured in households' maize stored for home consumption, P_{sj} is a vector of group level indicator variables indicating the matched pairs, ε_{ij1} is the error term and ε_{j1} is the group level error term. Standard errors will be estimated using a method appropriate to small numbers of clusters as described by Cameron, Gelbach and Miller (2008).

To test Hypothesis 1, we will test whether $\beta_1 = 0$.

6.2 Adoption of Aflasafe

Adoption and intensity of adoption will be analyzed in two ways: 1) considering only those farmers (up to 6 per group) from whom individual-level survey data were collected; and 2) using data from all members listed during the initial group census. For each specification, we will estimate both a parsimonious model (treatment indicators only) and a model including baseline control variables. For the specification including all group members, means of control variables measured through the individual baseline survey will be used.

Hypothesis 2

To assess the effect of the premium market linkage treatment on farmers' adoption of Aflasafe, the following two equations will be estimated (with and without controls), using the two samples (baseline survey respondents and all group members) as described above. In all cases, the sample will be limited to farmers in groups assigned to the input linkage treatment.

At the individual level, we will estimate:

$$Adoption_{ijv} = \alpha_2 + \beta_2 \cdot Market_v + (\gamma_2 \cdot X_{ij2}) + \varepsilon_{ijv2} + \varepsilon_{v2}, \qquad (6.2)$$

Where $Adoption_{ij}$ captures adoption or adoption intensity of Aflasafe by farmer *i* in farmer group *j*, $Market_j$ indicates whether the group was assigned to the output linkage treatment, and X_{ij} is the vector of individual- and group-level controls measured through the baseline individual and group-level surveys respectively, as defined in Section 3, plus an indicator for assignment of the group to each bundled money-back guarantee cluster. ε_{ij2ca} is the individual level error term and ε_{c2a} is the village level error term. Standard errors will be estimated using the Huber and White's sandwich estimator to account for clustering at the village level.

To test hypothesis 2, we will test whether $\beta_2 = 0$.

Hypothesis 3

To assess the effect of the bundled insurance treatment on farmers' adoption of Aflasafe, the following equation will be estimated (with and without controls) using the two samples (baseline survey respondents and all group members) as described above. In all cases, the sample will be limited to farmers in groups assigned to the input linkage treatment. Each of the four specifications of this regression will be tested using two approaches to the estimation of clustered standard errors. The primary specification will be determined based on the nature of intra-cluster correlation.

1. Bootstrapping to account for a small number of treatment clusters

The first approach accounts for potential intra-cluster correlation of standard errors at the level of the geographical clusters defined for assignment of the bundling money-back guarantee treatment. This approach takes a conservative approach to the potential spatial correlation of unobservable variables and outcomes of interest. We will estimate the following equation:

$$Adoption_{ijk} = \alpha_{3a} + \beta_{3a} Bundled_j + (\gamma_{3a} \cdot X_{ij3a}) + \varepsilon_{ijk3a} + \varepsilon_{k3a}, \qquad (6.3.1)$$

Where $Adoption_{ij}$ captures adoption or adoption intensity of Aflasafe by farmer *i* in farmer group *j*, in money-back guarantee treatment cluster *k*, $Bundled_j$ indicates assignment to the bundled insurance treatment, and X_{ij3a} is a vector of control variables, to be determined as follows. All individual and group level control variables listed in Section 5, plus the market linkage treatment, will be included in a regression specified as above, but omitting the final (cluster-level) error term. The control variable in this regression with the highest p-value will be dropped, then the regression will be re-run with the smaller number of controls. Again, the control variable with the highest p-value will be dropped, and so on, until only the treatment indicator plus three control variables remain. ε_{ijk3a} is the individual level error term and ε_{k3a} is the group level error term. For this equation, we will estimate standard errors using a method appropriate to small numbers of clusters as described by Cameron, Gelbach and Miller (2008).

2. Clustering at the farmer group level

The second approach is appropriate if the degree of spatial correlation of standard errors is less important than the precision gained by controlling for additional baseline variables. We will estimate the following equation:

$$Adoption_{ijvk} = \alpha_{3b} + \beta_{3b} Bundled_j + (\gamma_{3b} \cdot X_{ijv3b}) + \varepsilon_{ijk3b} + \varepsilon_{v3b}, \qquad (6.3.2)$$

Where ε_{j3b} is a standard error term defined at the village level, and X_{ij3b} is a vector including the output market linkage treatment and all individual and group-level control variables listed in Section 5. Standard errors will be estimated using the Huber-White sandwich estimator to account for clustering at the village level.

To test hypothesis 3, we will test $\beta_{3a} = 0$ and $\beta_{3b} = 0$

6.3 Relative level of aflatoxin contamination

Hypothesis 4

The following regression will be estimated to test hypothesis 4, limiting the sample to farmers within groups assigned to the input market linkage treatment:

$$Safety_{cijv} - Safety_{sijv} = \alpha_4 + \beta_4 \cdot Market_v + (\gamma_4 \cdot X_{ij2}) + \varepsilon_{ijv4} + \varepsilon_{v4}, \qquad (6.4)$$

Where $Safety_{ci}$ is the level of Aflasafe used, or the negative value of aflatoxin contamination measured, in a sample of maize taken from that stored for consumption by household i; $Safety_{si}$ is the level of Aflasafe used, or the negative value of aflatoxin contamination measured, in a sample of maize stored for sale by household i; ε_{ij4} and ε_{j4} are individual and village-level error terms respectively, and $Market_j$ and X_{ij2} are defined as in equation 6.2. Standard errors will be estimated using the Huber-White sandwich estimator to account for clustering at the village level.

To test hypothesis 4, we will test whether $\beta_4 = 0$.

7. Heterogeneous effects

7.1 Heterogeneous to be tested

The following individual-level heterogeneous effects will be tested:

7.1.1 Take-up of insurance within the optional money back guarantee treatment is higher for risk averse individuals

7.1.2 The effect of bundling the money back guarantee with Aflasafe on Aflasafe adoption is stronger for risk averse individuals

7.1.3 Take-up of insurance within the optional money back guarantee treatment is higher for individuals that are more familiar with insurance

7.1.4 The effect of bundling the money back guarantee with Aflasafe on Aflasafe adoption is stronger for individuals that have experience with insurance and have received a payout before

7.1.5 Impacts of the market linkage treatment are stronger for individuals who sold maize in the previous season

7.1.6 The input linkage treatment effects are higher for people with higher levels of baseline knowledge about aflatoxin and the potential harmful effects of consuming contaminated maize

The following group-level heterogeneous effects will be tested:

7.1.7 The output market treatment effects are stronger for the groups characterized by higher levels of trust

7.1.8 The output market treatment effects are stronger for the groups with higher levels of capacity

7.2 Estimation strategy

Heterogeneous effects will be tested as specified above while adding the necessary interaction terms, except that to preserve degrees of freedom, additional variables will be dropped from regressions testing for heterogeneous effects in the impact of the bundled money-back guarantee treatment in the modified equation 6.3.1 (accounting for the small number of clusters).

8 Other issues to account for during analysis

8.1 Attrition

Attrition at the individual and group level will be calculated after all the data is collected and tested for balance across treatments. This will be done for each of the hypothesis tests shown above by regressing attrition on the treatment dummies and clustering the standard errors as in the hypothesis test. If attrition is unbalanced, the estimation strategy will be adjusted to account for this nonrandom attrition using Lee bounds.

8.2 Spillovers

Both informational spillovers and our inability to exclude farmers from accessing either Aflasafe or the market premium through neighboring farmer groups could affect the estimated treatment effects. For example, farmers in the non-market linkage groups could sell maize through groups that are offered the premium. Farmers in the bundled money-back guarantee group could potentially buy Aflasafe without the guarantee through groups in the optional treatment. Negative impacts on adoption could arise through reference effects, for example if farmers in the bundled money-back guarantee treatments hear that those in the optional money-back guarantee are able to purchase Aflasafe at a lower cost.

We will ask endline survey respondents whether they purchased Aflasafe or sold maize through a group other than the one through which we identified them, and whether they know of other farmers who were offered an opportunity to purchase Aflasafe, and under what conditions.

We will also test for potential spillover effects by estimating regressions as specified above that also include a dummy variable indicating proximity to a farmer group (or individual farmer) in the treatment group through which spillover effects could potentially occur. If this variable is significant or affects the magnitude of the estimated treatment effect substantively, this will be taken as evidence of a spillover effect. A range of proximity thresholds will be tested to characterize the nature of spillovers.