Farmer experimentation as a public good. Experimental evidence on approaches to facilitate learning about new maize seed varieties in Kenya

Michael Ndegwa*
International Maize and Wheat Improvement Center (CIMMYT)
M.Ndegwa@cgiar.org

Sarah Kariuki
International Maize and Wheat Improvement Center (CIMMYT)
S.W.Kariuki@cgiar.org

Hope Michelson
Department of Agricultural and Consumer Economics, University of Illinois
Hopecm@illinois.edu

Mercy Mbugua
International Maize and Wheat Improvement Center (CIMMYT)
M.Mbugua@cgiar.org

Annemie Maertens
Department of Economics, University of Sussex
A.Maertens@sussex.ac.uk

Jordan Chamberlin
International Maize and Wheat Improvement Center (CIMMYT)
J.Chamberlin@cgiar.org

Jason Donovan
International Maize and Wheat Improvement Center (CIMMYT)
J.Donovan@cgiar.org

* Corresponding author

Abstract

Farmer experimentation with new technologies is fundamental to adoption, disadoption, innovation, and climate adaptation. Even so, this experimentation often takes place in semi-private ways that limit the positive externalities of farmer effort, potentially slowing learning by others. We implement a village-level randomized controlled trial to evaluate the effects of three scalable strategies to promote farmer experimentation and learning. Across 156 villages in Kenya, we vary the proportion of farmers to whom we offer a new hybrid maize seed; we additionally split the low concentration treatment arm to an incentive and non-incentive group. In one third of the treatment villages, 10% of villagers receive 0.5 kgs of maize. In a second treatment arm, 10% of the farmers received 0.5 kgs of maize plus incentives to compensate them for the knowledge that they generate and share based on their experimentation. In the third treatment arm a considerably higher share - 35% of the farmers - receive the maize seed. All participant farmers receive a sign to post next to their experimental field to mark their plot and their work.

We study the relative effects of these treatment arms on farmer experimentation, learning, perceptions and yield expectations, and adoption of new maize varieties. Our design with 4,160 farmers will test the effects of different levels of saturation within a village against a model where farmers are compensated for their experimental efforts.

Keywords: risk, agriculture, climate, technology adoption, poverty

JEL codes: D81; D25; O12; O13; Q12; Q54

1. Introduction

Farmers experiment with new technologies, new crops, new markets. Sometimes they experiment on a small plot within a larger field. Sometimes they quit the technology they are trying out. Sometimes they adopt a new crop or variety with sufficient scale or success that others notice and it spreads quickly and decisively across a region (Griliches 1957). Sometimes farmers experiment with a new technology but it fails to take hold, despite its demonstrated success in other, similar regions (Gollin et al., 2021).

Much farmer experimentation is presumably private and small scale, with private costs in farmer time and effort and in land and inputs allocated away from other crops. Experimentation generates private learning and benefit to the farmer. It can also generate positive externalities within a village. How big those positive externalities are will depend on how public and visible the farmer's experimentation is and on how much they share both the process of what they are doing and what they learn.

It is not a priori clear what the optimal amount of experimentation is either for an individual farmer or for a village as a whole. Farmers may under-provide knowledge gleaned from their experiments if they view their efforts as exclusively (or primarily) generating a private benefit. But observing the outcomes of many farmers' experimentation in a given growing season may be just as powerful and informative as observing the process and yields of a smaller number of experimenting farmers being compensated for their efforts. This paper uses a randomized controlled trial to study the effects of experimental scale (the share of experimenters in a village) and experimenting incentives on farmer learning and adoption. The technology we study is improved hybrid maize varieties. We site our research in Kenya, a country with numerous hybrid maize choices but very low varietal turnover by farmers; many report having grown the same hybrid maize variety for decades, even as conditions and options have evolved significantly (Rutsaert and Donovan, 2020).

Across 156 Kenyan villages, we vary the share of farmers to whom we offer a new hybrid maize seed. We further split the low concentration treatment arm to an incentive and non-incentive groups. Our three treatment arms therefore include a low concentration treatment in which 10% of villagers receive seed, a low concentration plus incentives treatment in which hosts are compensated for participation in the experiment and for sharing information actively and intentionally with their fellow villagers, and a high concentration treatment in which 35% of villagers receive the seed. All seed recipients receive a sign to mark their experimental plot and to indicate their participation in the experiment.

It is not clear that more is better when it comes to learning. In fact, it may be possible in this setting for an individual or a village to have *too much* experimentation and we design our experiment with attention to this possibility. Traditional Bayesian models of learning assume more information is always better, at least if the goal is learning the true signal. Although there are diminishing returns to this learning (so the tenth signal is less informative than the ninth), the marginal contribution is positive. But each experiment comes with a cost; at some stage the additional value of information may fail to exceed the marginal cost of producing it. Those calculations change fundamentally of course if the farmer includes the social benefits their experimentation generates for other farmers.

At the village level, considerable research has focused on the optimal way to seed information and technologies to promote farmer learning and adoption (Beaman et al. 2021; Beaman and Dillon, 2018; Bandiera et al. 2022; Kempe et al, 2005; Hinz et al., 2011). But less attention has focused on the fact that scale could have countervailing positive and negative effects at the village level.[1] For example, more observations of a new technology in a given season might not be better if heterogenous growing conditions within a village contribute to different yield outcomes from the same technology. If individuals focus on the negatives – on low yield realizations – more observations could deter adoption but may provide farmers more complete information about heterogeneity in returns. Village farmers may be more likely to observe a negative outcome, especially within a village characterized by significant heterogeneity in soil quality, growing conditions, or farmer ability. If farmers anchor on these lower tail yields or if they compare themselves to farmers with low yields, they may fail to adopt technologies that would have benefitted them. Another possibility: more individuals adopting a technology (or demonstrating it) could create some pressure within a village which could be welfare reducing if there is heterogeneity in returns.

[1] For example, a woman we interviewed during focus group discussions complained that many farmers in her village were in the habit of trying a new maize seed every year. She couldn't tell what people were growing; she could not make comparisons; she could neither aggregate nor learn from what they were doing. This is a case of considerable heterogeneity in what farmers are experimenting with, distinct from our framing, which is about scale and incentives to share information about a single technology.

Our paper contributes to the literature on farmer learning and technology adoption. Economists have studied a range of frictions that can impede technological adoption and diffusion (De Janvry et al., 2017; Magruder, 2018, Udry and Suri 2022, Bandiera and Rasul, 2006, Conley and Udry, 2010, Munshi, 2004). One way that economists have worked in recent years to understand the relevant frictions is to experiment with understanding the tradeoffs among different models of introducing farmers to new technologies. Several papers have contrasted for example strategic seeding information within villages and social networks with what is known as broadcast diffusion (Matous 2023; Kelley et al. 2023; Hinz et al. 2011; Banerjee et al. 2019; Beaman et al. 2021); others have compared the effects of demonstration plots worked by community members with field days attended by outside villagers on learning and adoption (Maertens et al. 2021). This work has established the importance of both social and physical proximity to social learning (Kondylis et al., 2017 for example) and has found evidence of spillovers (Hörner et al., 2022) across farmers in communities with demonstration plots and field days.

A few studies have experimented with farmer incentives for information sharing among farmers, with decidedly mixed conclusions about their effects. Benyishay & Mobarak (2019) show that peer-to-peer learning can be augmented with incentives to the disseminating farmers, with peer farmers paid based on other farmers' gains in knowledge about the technology. Shikuku & Melesse, (2020) and Shikuku et al., (2019) show that social recognition incentives increase knowledge spillovers to other farmers in villages and also changed information networks of both disseminating farmers and their neighbors, performing better than private material incentives. Okello et al., (2023) found that social incentives reduced the likelihood of the trained progressive farmers reaching out to co-villagers to share information and discuss farming.

Our research is relevant to an important applied problem facing farmers in many low income countries: whether the pace of technical adaptation can adjust to the increasing pace of environmental change, especially among poor producers and farmers. Climate change makes the question of farmer experimentation and learning both more urgent and more challenging. As conditions change, farmers need to adapt, and will need to respond to shifting growing conditions. The returns to agricultural technologies are highly stochastic across years as they are largely determined by weather. Crop yields may become more variable in the coming years in rainfed systems, driven by increasing fluctuations within and across seasons in temperature and rainfall. This may make learning harder as the signal from the technology will become harder to detect within the inter-annual noise, limiting what a farmer can learn from a single or a handful of years.

Our results will provide new insights into why farmers are so reluctant to change varieties, and into how small farmers understand the risks and benefits of new technologies. Farmer uptake of new seed varieties is limited by both demand and supply in local input markets. Constraints to adoption include farmers' lack of familiarity with, interest in, and trust in new varieties. Farmer seed choices are sticky; many cultivate varieties they have grown for years (Hugo 2023; Smale and Olwande, 2014), despite extension efforts and agronomic demonstrations indicating that newer varieties perform better by some key metrics, especially having been bred with the current production challenges in mind. New varieties considered by breeders and agronomists to be optimal for a given area thus have had limited effective demand by farmers. Known varieties, even if likely to be inferior performers to new varieties, remain strongly preferred to new varieties.

2. Research Design

Outcomes of interest

- Adherence to experimental recommendations (for trial hosts in treated villages)
- Willingness to experiment with new varieties
- Village level social learning about experimental varieties and spillovers of information
- Perceptions about new varieties (vis-à-vis old known ones) on performance, quality and production costs
- Awareness and exposure to the specific varieties used in the study (experimental varieties)
- Varietal switching – from old well known to new promising varieties
    - Proportion of planted varieties that are new
    - Proportion of maize area cultivated with new varieties
    - Weighted average age of varieties planted by households
- Expected yield distributions for new varieties vis-à-vis established ones (status quo)

Hypotheses

Hypothesis 1: adherence by trial hosts to experimental recommendations, especially on dissemination of information, will differ across treatments arms.

Hypothesis 2: intervention will lead to improvements across the target outcomes among the treatment groups: T1|T2|T3>T0 where T0=control group

Hypothesis 3: outcomes will differ across the first treatment arm (T1) and the other two treatment arms (T2 and T3): Test T1<T2|T3

Hypothesis 4: outcomes will differ across T2 and T3, but we have no priori expectation of the direction of the differential effects: Test T2=T3

Hypothesis 5: outcomes will vary across the host farmers **(n1)** (those directly treated/received and planted trial-packs) and the non-host farmers **(n2)** (the survey sample who did not receive trial-packs but are used to assess social learning and information spillovers): where n1>n

Basic methodological framework / identification strategy

We take an experimental approach to this question where we vary the proportion of farmers in a village to whom seed trial packs are offered, splitting the low concentration treatment arm to an incentive and non-incentive group:

- low concentration treatment: 10% of villagers receiving trial packs.
- low concentration plus incentives treatment in which hosts will be compensated for participation, to evaluate the role of incentivizing host farmers to share information actively and intentionally with their fellow villagers.

- high concentration treatment: 35% of villagers receiving trial packs.

Our design will inform decisions concerning the proportion of farmers in a given area/region seed companies and other stakeholders should be distributing seed trial packs to and whether incentives are necessary to promote experimentation. These are important decisions given that scale directly affects marketing budgets for new varieties. Our work will measure the impacts of alternative promotion models on outcomes of interest at the farmer- and village-level, including awareness of new varieties, perceptions and beliefs about new varieties, demand and willingness to pay for new varieties.

Intervention

Our treatments vary at cluster/village level. Treatments are defined at the village-level in order to evaluate village-level effects on awareness, perception and demand for experimental varieties. Randomly selected households from the treatment villages will be presented with trial packets of two preselected experimental maize varieties for them to endogenously select one. The selected varieties must meet the following characteristics:

1. they are relatively new varieties – at most 5-years and below since they were commercially released
2. they are preferred varieties for that locale, as recommended by breeders and agronomists working in that geographic area,
3. they are available in local markets BUT
4. are not yet widely used by local farmers (purchased by <10% of buyers?).

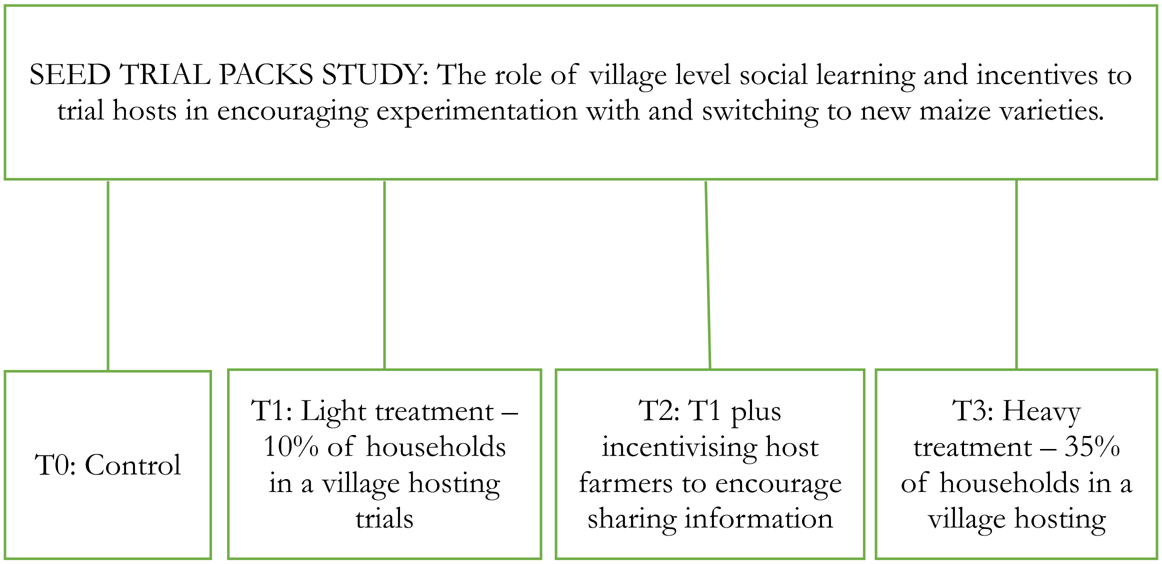
Our experimental design includes a control group and three treatment arms as defined below:

- T1: Light treatment – this involves low concentration in terms of the proportion of farmers hosting the trials in a village. We plan to reach a 10% concentration level for this treatment arm. The randomly selected hosts will be given 0.5kg of seeds to experiment with.

- T2: Light treatment + incentives – we layer incentives to host farmers on top of T1. The hypothesis here is that incentivizing host farmers could encourage them to share information among their counterparts, which could then improve the effectiveness of trial packs model, even with low concentration. For the incentives, we propose Ksh 1500 sent to host farmers under this treatment arm in three instalments of equal amount and intervals. This will be coupled with follow-up calls to discuss the progress of their trial plots and their engagement with fellow villagers on the trials and varieties planted.

- T3: Heavy treatment – a high concentration of trial hosts in a village is envisaged here. We plan with 35% of households in a village hosting the trials and 0.5kg of seeds for each host farmer.

- T0: Control – this is a comparison group where no intervention will be implemented.



Within each treatment village, the 0.5kg seed packs will be distributed to households randomly selected from village level sampling frames, where every household in a village has equal chance of being selected as a trial host. The lottery-based distribution, coupled with random selection of households to the study, should partially alleviate concerns about potential negative impacts of freely distributed goods on subjective valuation (and consequent diminished willingness to pay for seed in post-experimental periods), because while the seed is freely distributed to lottery winners, it is not freely distributed to everyone in the community.

Seed distribution will be conducted at the same time as baseline survey and will be accompanied by an information campaign. The campaign will be uniform across all treatments and consist of a village/cluster level meeting for all the selected trial hosts. Trained enumerators will conduct training and sensitization sessions at the meeting describing the purpose of the experiment (i.e., to facilitate village-level learning about new varieties) and norms for experimentation (e.g., how to set up experimental plots on farms, controlling for as many factors as possible, in order to compare outcomes across different varieties). All seed pack distributions will be accompanied by signs which the farmer will be encouraged to display in the experimental plots she/he establishes with the trial packs. The implementation of such experimental norms and signage will not be enforced but will be monitored and recorded as part of the experiment's data collection activities.

Sample and statistical power

We have selected four Kenyan counties for the implementation of this study: Embu, Kirinyaga, Vihiga and Busia counties. We also obtained a Kenya villages map with their names and GIS coordinates and extracted villages for the four counties. After some minor cleaning to remove duplicates, we were left with 211 villages in Embu County, 230 in Kirinyaga, 294 in Vihiga and 256 in Busia. After excluding villages in urban, semi-urban and in forests areas, we have randomly selected 52 villages per county and 208 in total. We hence propose a random administration of 208 villages into 3 treatments and one control group, all of equal size (i.e., 52 villages in each group). Since we do not have data on outcomes of interest in our target population, our power calculations require assumptions. Figure 1 and 2 below show the detectable effect size for pairwise comparisons of different numbers of control and treatment groups. If we want to be able to compare each treatment against the control, as well as compare treatments against one another, then we maximize power by making each group the same size. This means that with a total of 208 candidate villages, we can work with 4 groups (1 control and 3 treatment arms) of 52 villages each. If we want to use the likelihood of purchasing a new variety in any given year as the basis for power analysis, we have no data on this, so we must assume some distributional characteristics.

To begin with, let us assume that 10% of HHs are switchers, or do purchase a new maize variety in any given year. With a fixed number of clusters, we then examine the number of households required for data collection in each village to detect a treatment-induced change from that 10%, assuming that the intra-cluster correlation is 0.2 and standard levels of acceptable type I and II errors (i.e., alpha = 0.05 and power of 80%, using a one-sided test).

We first consider an analysis of spillover effect among non-hosts that compares T1 (the lightest treatment) to the control group. Assuming baseline/control outcome of 10% switchers and MDE of 0.086 or 8.6 percentage points increase in uptake of new varieties among the treatment villages, we need 16 households per village. For the analysis of trial-hosts outcomes, one can assume a higher MDE since they receive the treatment directly (heavier treatment). With 52 clusters and MDE of 0.112 or 11.2 percentage points increase in uptake of new varieties among the hosts in treatment villages, we need 4 host households interviewed per village. These results are shown in Figure 1 below.

For an analysis that compares treatment groups among each other, we assume T1 as the comparison group for T2 and T3, and assume that, after the intervention, 20% of T1 will be switchers. Assuming equal sample size across experimental groups, we shall be able to detect an MDE of 0.106 (10.6 percentage points) with 16 non-host households surveyed and 0.137 (13.7 percentage points) with 4 hosts households surveyed. These results are shown in Figure 2 below.

Figure 1: detectable effect size from assumed 10% baseline share of farmers buying new varieties in any given year

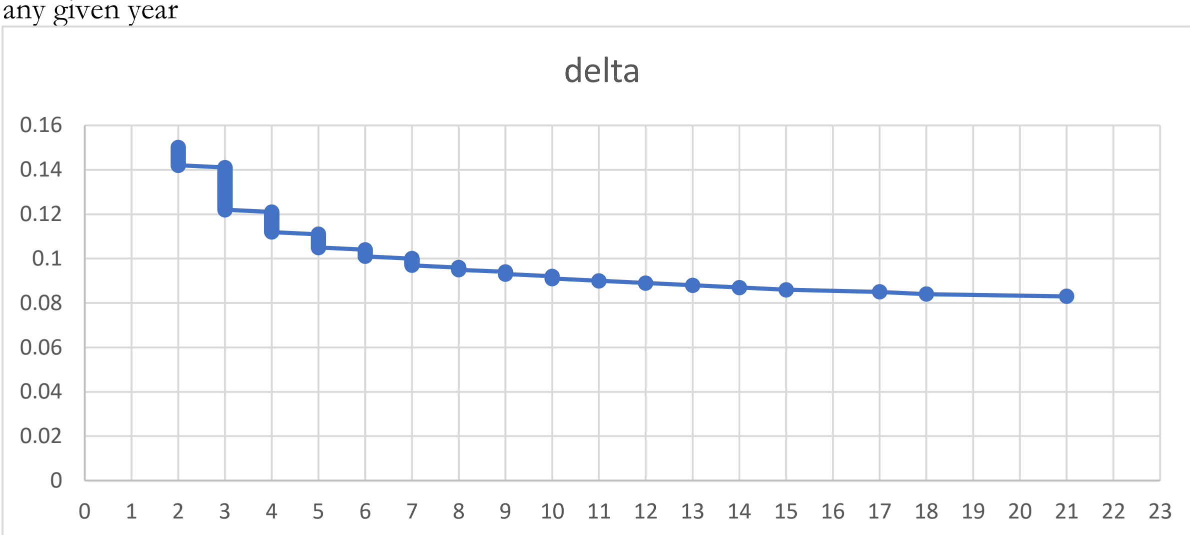
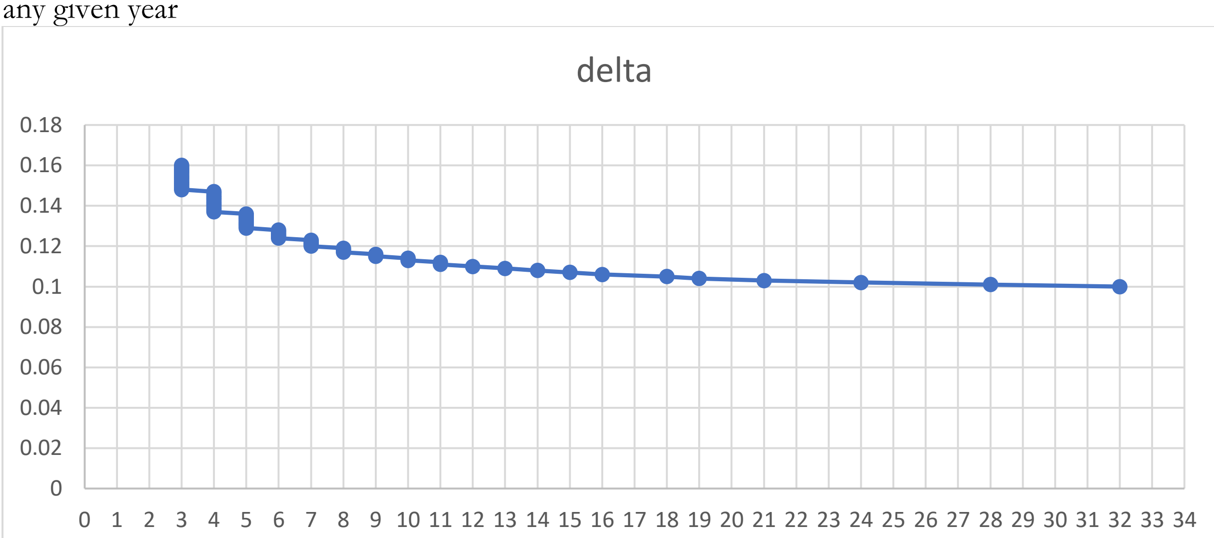


Figure 2: detectable effect size from assumed 20% baseline share of farmers buying new varieties in any given year



3. Data

Data collection and processing

Data for this study come from two main sources:

1. Three period panel of face-to-face household surveys targeting 4,160 households. Out the total sample, 1,040 are from control villages, 2,496 are from treatment villages but did not

receive trial-packs (were not directly treated) while 624 are from treatment villages and received trial packs (were directly treated). The three-period panel will consist of:

- Baseline survey which was implemented jointly with trial packs distribution in February-March 2023. The baseline survey collected household demographic and socio-economic information which will be necessary for characterizing our sample, checking for treatments randomization effectiveness through balance checks and controlling for households' heterogeneity at baseline in our regression estimations to increase precision. We also collected information on our key outcomes which will we intend to include in our regression estimations as controls to increase precision.
- Midline follow-up survey which will be implemented in October-November 2023, or right after the seed-purchasing and planting period for the second season is over. All the study areas experience bimodal rainfall season with long rains season coming first between March – July and short rains season coming second between October – December. While the interventions are implemented during the long rains season, we plan to track our outcomes at the two subsequent seasons. Our interventions are more about behavioral outcomes and less about promotion of particular hybrids and we want to assess if this will affect the farmers' behavior and choices during the short rains season.
- Endline follow-up survey which will be implemented in April-May 2024 or one year after the interventions were implemented. This will be the main follow-up survey and unlike the mid-line survey, it will help assess the outcomes of interest for a season similar to one that the project interventions were implemented. The follow-up surveys, both midline and endline, will collect data on outcome measures that we listed above.

2. Three round phone surveys implemented among all the trial hosts during the March-August cropping season which is the intervention season for this study. The calls started at about three weeks after planting period with two consecutive calls scheduled at about mid the season and a few weeks after harvesting period. During the first calls, we collected basic household demographics and socio-economic information which will be used to characterize the all the hosts, check for randomization balance and control for host related outcome estimation. We also collected information regarding trial set up compliance including whether they planted or not, planting date and an assessment of how they treated trial plots in reference to their other maize fields – regarding plot selection, seeding and fertilizer application rate, intercropping, irrigation, etc. During all the three call rounds, we ask them the number of farmers they have reached out to or have shared information about the trials with since the beginning of the season and in the last past one week.

Besides the two main sources of quantitative data above, we shall also be drawing insights from a pre-RCT qualitative scoping work which included 8 FGDs with farmers and several KIIs with agro-dealers. The farmer FGDs were particularly helpful in finalizing the study design and the questionnaires while the KIIs with agro-dealers were particularly helpful in selection of maize varieties to be used for the study.

Variations from the intended sample size

This study has enrolled 20 farmers in 208 clusters (totaling 4,160 farmers (624 trial hosts and 3,536 non-hosts)) for three rounds of household surveys which are roughly a cropping season a part. They all were surveyed at baseline, with minimal replacements for households initially listed but were either unavailable or unwilling to participate in the study. The study has also enlisted a separate random sample of 2,132, trial hosts with varying numbers per cluster depending on the treatment arm for respective clusters, who are taking short surveys over the phone. There are two main threats to maintaining this sample size namely attrition and imperfect compliance.

To minimize attrition, we collected detailed contacts for all participants, including alternative contact numbers in case we miss them either at home or on the primary contact numbers. We also work with village elders who act as guides and help with identification and tracking of farmers. We will also make repeated interviews and book appointments with those who may be initially busy. Since the follow-up tool will be substantially short, we shall also conduct phone interviews as a last resort for those who are willing to participate but are either not available during the survey period or may have moved out of the study areas. We are also implementing this study in-house and we shall endeavor to maintain the same survey teams throughout the study as much as possible. Previous experiments in similar contexts have reported attrition rates of less than 10% and we also anticipate low attrition in this study as well.

Imperfect compliance could arise due to either 1) a selected host not receiving the trial packs either due to implementations hurdles or non-willingness to participate or 2) a host farmer receiving the trial pack but not setting up experimental plots and hence denying their village mates an opportunity to observe and learn. To reduce non-compliance resulting from the former cause, farmers selected as hosts were required to come to a central place to collect their pack and get some induction into the protocol and those who could not but were willing were delivered to at home by the survey team. We, however, had few cases of non-willingness of which were replaced on the spot with randomly selected fellow villages. For the second cause of non-compliance, we made every effort to confirm that the selected farmers were maize farmers and that they were willing and able to set aside a plot enough for 500 grams of trial seeds. During the first round of phone surveys, we asked them if they planted the trial packs and only 0.7% indicated that they did not plant. We shall compute a village level compliance rate variable which we shall use to control for this in our main estimation models.

4. Analysis

We shall mainly rely on simple means difference (SMD) estimated with Ordinary Least Squares (OLS) and Linear Probability Models (LPM) to estimate the effect of our treatments on the continuous and binary (respectively) outcomes listed above. For robustness and consistency check of our SMD results, we shall also apply analysis of covariance (ANCOVA), and fixed effects (FE) and present the results side by side.

Hypothesis 1: adherence by trial hosts to experimental recommendations will differ across treatments arms.

Adherence/compliance to the recommendations will be measured in three different ways: 1) Planted the trials which will be a binary variable equal to one if one planted the trial seeds and zero if one did not, 2) protocol followed which will be an index variable computed from a variety of questions which seeks to understand if the farmer treated the trial plots differently from the other maize fields. This includes planting time, fertilizer application rate, seeding rate, intercropping, plot selection, and irrigation, and 3) self-reported dissemination efforts which will be a discrete variable comprising of the number of farmers/other people a host reports to have reached at three different points within the season. Data for this hypothesis is being collected using three-rounds phone surveys among the trial hosts.

We expect that adherence to the recommendations the host farmers were given during seed distribution will differ across the three treatment arms. The first treatment, low saturation group, is the most basic treatment and hence we expect higher compliance among farmers in T2 and T3 in response to the "heavier" nature of the treatments. We believe the incentives in T2 and the increased peer-to-peer influence and competition in high saturation clusters of T3 should produce differential effect from T1. We will use regression equation 1 to compare adherence across treatment arms:

$$\forall_{i,j} = \alpha + \beta_1 T2_{ij} + \beta_2 T3_{ij} + \sum_{i=1}^{K} \beta_k X_{i0} + \pi_t + \varepsilon_j + \varepsilon_{ij} \tag{1}$$

Where $\forall_{i,j}$ represents adherence for household i in village/cluster j, $T2_j$ and $T3_j$ are village/cluster level treatment indicators equal to one if a household was randomly assigned to the respective treatment group (as defined above) and zero if in T1, X_{i0} represents a vector of farmers baseline socio-economic characteristics, π_t is the panel period indicator equal to zero at baseline, one at midline and two at endline, ε_j is the cluster level error term and $\varepsilon_{i,j}$ is the farmer/individual idiosyncratic error term. The parameters of interest are β_1 and β_2 which capture the treatment effects. These coefficients will reveal whether T2 and T3 effects were substantially and significantly different from T1 effects. We shall then use coefficient comparison approach to explore whether T2 and T3 are significantly different.

Hypothesis 2: intervention will lead to improvements on the target outcomes among the treatment groups: T1|T2|T3>T0 where T0=control group… *add cluster level percentage of compliers in the model*

In reference to the control group, we expect our interventions to yield some benefits among the treated villages and households in regard to awareness and perception of new maize varieties as well as experimentation and uptake of newer varieties. We will use regression equation 2 to estimate the effects of our treatments on a set of target outcomes:

$$Y_{i,j} = \alpha + \beta_1 T1_{ij} + \beta_2 T2_{ij} + \beta_3 T3_{ij} + \beta_4 Y_{ij,0} + \sum_{i=1}^{K} \beta_k X_{i,0}^{k} + \pi_t + \varepsilon_j + \varepsilon_{ij} \tag{2}$$

Where $Y_{i,j}$ is a vector of the observed outcome variables for household i in village/cluster j post treatment, $T1_j$, $T2_j$ and $T3_j$ are village/cluster level treatment indicators equal to one if a village was randomly assigned to the respective treatment group (as defined above) and zero otherwise, π_t is the panel period indicator equal to zero at baseline, one at midline and two at endline survey, and $\varepsilon_{i,j}$ is the idiosyncratic error term for household i in village j. To enhance precision in our estimation, we also include $Y_{ij,0}$ which represents the baseline measure of the outcome variable and $X_{i,0}^{k}$ which represents baseline socioeconomic covariates chosen through double post-lasso procedure as defined by Belloni et al. (2014), The parameters of interest are β_1, β_2 and β_3 which respectively capture the treatment effects for the three treatment arms.

Hypothesis 3: outcomes will differ across the first treatment arm (T1) and the other two treatment arms (T2 and T3): Test T1<T2|T3

We expect that the heavier treatments (T2 and T3) will yield higher benefits than the most basic treatment arm namely low saturation (T1). We believe the host farmers in T2 who received a monetary incentive to encourage them to comply to the protocol will put more effort in disseminating information about the new varieties and experimentation in general. We will use the coefficients comparison approach to test this hypothesis, where, from Equation 2, we will examine if β_2 is substantially and significantly greater than β_1:

$$\beta_1 T1_{ij} < \beta_2 T2_{ij}$$

Similarly, the higher saturation in terms of proportion of farmers planting trial plots in T3 villages should lead to higher effects compared to low saturation in T1. The expect the differential effect to come from the fact that non-host farmers in T3 have more sources to learn from and be influenced by which should increase learning and awareness of new varieties as well as improve their perception towards them and eventually encourage experimentation with and adoption of the new varieties. With higher saturation, besides an increase in absolute number of learning points, it increases the chances of receiving information from a source (someone) that is trusted by the recipient. To test this hypothesis, we will examine if β_3 in Equation 2 is substantially and significantly greater than β_1:

$$\beta_1 T1_{ij} \; < \; \beta_3 T3_{ij}$$

However, recent literature on social networks and social learning indicates that it is not always that higher saturation will lead to greater effects. Some of the pitfalls associated with high saturation include: 1) conflicting information, 2) too much information, 3) low motivation by disseminators due to loss of "teacher" privileged title when it's given to almost everyone, among others.

Hypothesis 4: outcomes will differ across T2 and T3: Test T2=T3

We hypothesize that the effects of low saturation with incentives treatment arm (T2) will be significantly different from the high saturation treatment arm (T3) but we have no priori expectation of the direction of the differential effects. We hence explore the appropriate strategy to achieve scale: is it low saturation combined with incentives or just high saturation. To test this hypothesis, we will examine if β_3 in Equation 2 is substantially and significantly different than β_2:

$$\beta_1 T2_{ij} = \beta_3 T3_{ij}$$

Hypothesis 5: outcomes will vary across the host farmers **(n1)** and the non-host farmers **(n2)**

The study design consists of farmers who were directly treated within treatment villages and a random sample from the same villages who did not receive the trial packs but are used to assess the spillovers. We anticipate substantially larger effects on those who are directly treated (n1) compared to their village mates who did not receive trial packs (n2). This will be estimated from total sample of 3,120 farmers comprising of 624 who were directly treated and 2,496 who did not receive the pack but from treatment villages. We will use equation 3 for this estimation:

$$y_{i,j} = \alpha + \beta_1 H_{ij} + \pi_t + \beta_4 Y_{ij,0} + \sum_{i=1}^{K} \beta_k X_{i,0}^{k} + \varepsilon_j + \varepsilon_{ij} \tag{3}$$

where, in addition to already defined abbreviations, H_{ij} is a binary indicator equal to one if a farmer was directly treated (trial host) and zero if did not directly receive the treatment (non-host).

Assumptions

- The identified maize varieties are (a) likely to outperform the commonly used varieties, (b) are available from local stockists, and (c) are not currently being used by local farmers in appreciable quantities.
- The distributed experimental varieties are not among the main varieties planted by a farmer in the non-experimental plots – for comparison's sake by the host and village mates.
- The farmers' status quo (established) maize varieties are a sufficient control to enable observational comparisons by farmers in the village

Procedures for dealing with attrition, missing values and outliers

To deal with attrition at mid- and end-line surveys, we will: 1) make extra efforts as mentioned above to reach almost all the respondents, 2) test for non-random attrition or check for attrition bias in our sample. We will generate a binary indicator for each follow-up round (midline and endline) equal to one if a household dropped from the study at the respective survey round and 0 otherwise. We will then estimate a series of probit models with this indicator as the dependent variable to analyze association between attrition and (i) random assignment to control and treatment arms, (ii) outcome variables, and (iii) socioeconomic control variables for the baseline sample. This will be done separately for each experimental group and then for the pooled/overall sample, 3) use bounds approach to examine the robustness of our results to attrition, if any (Lee, 2009; Tauchmann, 2014).

We were able to reach our target sample size at baseline, albeit with minimal replacements as explained above. While analyzing the RCT, we shall dummy out any missing baseline data for the selected covariates. We will create dummy variables for all covariates with missing data which will be equal to one if missing and zero if non missing, and then replace the missing values with zero across all the baseline covariates, and include both the original variables and their respective dummies as controls in our estimation models.

To deal with large outliers, we will winsorize our continuous outcome variables at 99^{th} percentile.

Multiple outcome and multiple hypothesis testing

In this study, we estimate the treatment effects of three treatment arms on several outcomes translating to multiple hypotheses tests. This raises the false discovery rate (FDR) concern where significant coefficients may emerge by chance when there are a large number of measured outcomes and tested hypotheses, even when there are no true treatment effects on the outcomes. To check for robustness of our results to this potential biasness, we will conduct multiple hypotheses correction tests using sharpened q-values (Anderson, 2008; Benjamini et al., 2006). Further, following Young (2019), we will conduct an F-test for all our outcome regression estimations to test the null hypothesis that the joint effect of the three treatments is zero.

Heterogeneous Effects

We have collected baseline information on households' demographic and socio-economic characteristics which we shall use to check for heterogeneity in treatment effects. As Chernozhukov et al (2023) advises, we are not restricting ourselves to examine treatment effects heterogeneity to a limited number of pre-determined subgroups as that amounts to throwing away a large amount of potentially valuable information. Instead, we aim to use the data to discover ex-post whether there is any relevant heterogeneity in treatment effect by covariates. However, to avoid overfitting, we shall use their (Chernozhukov et al., 2023) generic machine learning approach for predicting and making inference on heterogeneous treatment effects.

Bibliography

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, *103*(484), 1481–1495. https://doi.org/10.1198/016214508000000841

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, *28*(2), 29–50. https://doi.org/10.1257/JEP.28.2.29

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, *93*(3), 491–507. https://doi.org/10.1093/BIOMET/93.3.491

Benyishay, A., & Mobarak, A. M. (2019). Social learning and incentives for experimentation and communication. *Review of Economic Studies*, *86*(3), 976–1009. https://doi.org/10.1093/RESTUD/RDY039

Chernozhukov, V., Demirer, M., Duflo, E., & Fernández-Val, I. (2023). *Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India* (No. 24678; NBER Working Papers).

Hörner, D., Bouguen, A., Frölich, M., & Wollni, M. (2022). Knowledge and Adoption of Complex Agricultural Technologies: Evidence from an Extension Experiment. *The World Bank Economic Review*, *36*(1), 68–90. https://doi.org/10.1093/WBER/LHAB025

Kondylis, F., Mueller, V., & Zhu, J. (2017). Seeing is believing? Evidence from an extension network experiment. *Journal of Development Economics*, *125*, 1–20. https://doi.org/10.1016/J.JDEVECO.2016.10.004

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies*, *76*(3), 1071–1102. https://doi.org/10.1111/J.1467-937X.2009.00536.X

Matous, P. (2023). Male and stale? Questioning the role of "opinion leaders" in agricultural programs. *Agriculture and Human Values*, *1*, 1–16. https://doi.org/10.1007/S10460-023-10415-9/TABLES/3

Okello, J., Shikuku, K. M., Lagerkvist, C. J., Rommel, J., Jogo, W., Ojwang, S., Namanda, S., & Elungat, J. (2023). Social incentives as nudges for agricultural knowledge diffusion and willingness to pay for certified seeds: Experimental evidence from Uganda. *Food Policy*, *120*, 102506. https://doi.org/10.1016/J.FOODPOL.2023.102506

Shikuku, K. M., & Melesse, M. B. (2020). Networks, incentives and technology adoption: evidence from a randomised experiment in Uganda. *European Review of Agricultural Economics*, *47*(5), 1740–1775. https://doi.org/10.1093/ERAE/JBAA009

Shikuku, K. M., Pieters, J., Bulte, E., & Läderach, P. (2019). Incentives and the Diffusion of Agricultural Knowledge: Experimental Evidence from Northern Uganda. *American Journal of Agricultural Economics*, *101*(4), 1164–1180. https://doi.org/10.1093/AJAE/AAZ010

Tauchmann, H. (2014). Lee (2009) treatment-effect bounds for nonrandom sample selection. *The Stata Journal*, *14*(4), 884–894.

Young, A. (2019). Channeling Fisher: Randomization Tests and the Statistical Insignificance of

Seemingly Significant Experimental Results. *The Quarterly Journal of Economics*, *134*(2), 557–598. https://doi.org/10.1093/qje/qjy029