Can informed buyers improve goods quality? Experimental evidence from crop seeds

Eric Hsu[a] Anne Wambugu[b]

This version: October 2023

Abstract

We study whether adding informed buyers to a market can improve the quality of goods supplied by sellers, in an environment where goods quality is difficult to observe. To do so, we implement a market-level intervention, randomizing rural markets in Kenya into a community-wide information campaign. Small-scale farmers in treated market areas were trained to identify hybrid maize seeds that are quality-verified. In this setting there are widespread concerns about deceptive counterfeits and other uncertified seeds of lower quality. We find that observable markers predict seed quality. Treatment increased knowledge of these markers, affected seed purchase decisions, and increased maize production. Impacts were heterogeneous, with more educated and more remotely located farmers benefiting more. Uninformed buyers in treated communities did not gain better access to high quality seeds, as revealed by data from secret shoppers. These patterns can be explained by a model in which informed buyers can detect quality, and can change sellers if needed to obtain higher quality. However, more informed consumers hurts firm profits and can induce sellers to exit the market if upgrading quality is too costly. As a result, uninformed consumers may not benefit from the presence of more informed consumers. Consistent with these predictions, we find that treatment caused many seed sellers to exit the market. Taken together, the findings document new stylized facts and provide evidence relevant for boosting yields of a staple crop. More generally, they provide lessons concerning the role of improved consumer information in disciplining firms in low information environments.

[a] Yale University [b] Strathmore University. We thank Susanna Berkouwer, Stephanie Bonds, Ben Faber, Lauren Falcao Bergquist, Frederico Finan, Benjamin Handel, Jessica Hoel, Supreet Kaur, Kei Kawai, Samuel Leone, Jeremy Magruder, Aprajit Mahajan, Edward Miguel, Maximillian Mueller, Duncan Mwangi, Andrés Rodriguez-Clare, Joseph Wairiuko, Guo Xu, and participants at UC Berkeley Development Lunch, and IZA/FCDO G2LM-LIC Research Conference for feedback and helpful conversations. We thank Esther Kimani and Simon Maina at the Kenya Health Inspectorate Service, as well as Bright Simons, Carolyne Kimuhu, Timothy Maina, and Charles Ochieng from mPedigree for helpful conversations and their support for the project. We thank Noah Forougi, Cashua Wang, and Astri Doub for excellent research assistance. We are also grateful to REMIT Kenya staff – and particularly Carol Nekesa, Nettah Isavwa, Erick Emilio, Stephen Omare, and Christine Soita for their work on data collection and other field work for this project. The Agricultural Technology Adoption Initiative, IZA, FCDO, the Center for Effective Global Action, and the Private Enterprise for Development in Lower-Income Countries initiative provided financial support for the project, and we acknowledge that the views expressed are ours and not necessarily those of the above mentioned organizations and initiatives. Data collection was carried out under UCB CPHS protocol 2019-10-12666.

1 Introduction

When buyers cannot perfectly observe product quality, a market breakdown can occur in which offering high-quality products may not be profitable for sellers (Akerlof, 1970). A variety of mechanisms may help counteract such a market failure by providing a credible signal of product quality to consumers (Dranove and Jin, 2010). For example, firms may use branding or warranties to show they are incentivized to provide high quality. Government agencies can set quality standards and provide certifications or licenses to assure consumers of a certain level of quality in a market.

However, mechanisms to counteract the effects of limited consumer information may function imperfectly or be non-existent in many settings. Markets in low and middle income countries (LMICs), often have features that make them particularly vulnerable to problems with imperfect information. Lower access to information technology or lower levels of education may slow the diffusion of consumer information and hinder a firm's ability to build a reputation for providing high quality through word-of-mouth (Aker and Blumenstock, 2015). Policy solutions may be undermined by lower state capacity (Dal Bó and Finan, 2020). This could lead to less effective enforcement of quality certifications for example. Branding and reputation building mechanisms may also function less well if entry of deceptive counterfeits cannot be deterred through legal institutions.

In this paper, we study the effects of improving consumers' ability to detect quality in the hybrid maize (corn) seed market in rural Kenya. Several features of this setting may hinder the ability of buyers to know the quality of the product they buy (Bai, 2021; Bold et al., 2017; de Brauw and Kramer, 2022; Hoel et al., 2022). By nature, the quality of hybrid maize seeds is difficult to observe prior to purchasing and planting them. Even after planting, the performance of maize seeds depends on numerous factors, each of which can have a sizable impact on plant growth and quantities of harvested maize. As a result, maize yields tend to be noisy and give buyers a poor signal of seed quality. Typical crop planting timelines limit farmers to only observing one or two harvests per year. Therefore, infrequency of purchase may also limit consumer learning about seed quality. While regulations exist to establish minimum quality standards for seeds, enforcement is limited. For example, regulators conduct testing after manufacture prior to certification, but have very limited capacity to monitor quality along the entire distribution chain all the way to the end-user. As a result, there have been widespread concerns about the prevalence of seeds that are not quality verified in recent years, such as counterfeit, adulterated, and poorly stored seeds (Okinda, 2019).

Resolving information problems and improving market-supplied quality is likely to be particularly consequential in this setting. High quality maize seeds are a productive input that is critical for maintaining food security and battling chronic hunger and malnutrition in low-income areas of the world. Maize is a staple food in our study area in Western Kenya but also more broadly in Southern, Central, and Eastern Africa and in Central and South America. In these regions, over 20% of total calories are supplied by maize (Shiferaw et al., 2011). Our study area is in the East Africa region, where food security is a widespread and chronic issue and a major focus for policymakers. Over 30% of the population in East Africa experience severe food insecurity, lacking adequate access to enough safe and nutritious food for healthy development (FAO, 2020, 2021). This contributes to malnutrition as a major public health problem. For example, in 2019 an estimated 34.5% of under-age-five children in the region were stunted and 5.3% were moderately or severely wasted, with height-for-age or weight-for-age at least two standard deviations below the median in World Health Organization standards (FAO, 2020).

To evaluate the effects of improved consumer information, we conduct a field experiment in a sample of 386 rural market areas in Kenya. In randomly selected markets, we provided community-wide trainings to enable farmers to better detect quality-verified hybrid maize seeds. This market-level intervention makes it possible to identify effects of improved information on local market participants, including consumers who became informed due to treatment, uninformed consumers in the same communities, and seed sellers. We examine whether and how informed buyers can achieve improved agricultural outcomes. We also examine whether uninformed buyers in treated communities also experience benefits through improved seed quality. Using a series of surveys and secret shopper activities, we track outcomes for buyers and sellers (including knowledge, seed sales and purchases, and agricultural outcomes) for over one year after treatment. The inclusion of 104 pure control sites allows us to test whether baseline survey activities could have influenced outcomes. The extended follow-up period of over one year allows us to track possible convergence to a new equilibrium over the course of three planting seasons following treatment.

The analysis produces three main findings. First, the analysis shows that observable quality markers correlate with a key measure of seed quality, with packets missing one or more quality markers having lower lab-tested germination rates.[1] While germination rate is far from the only aspect of seed quality that is relevant for maize yields, this finding strongly suggests that agricultural gains should in theory be possible if farmers follow a simple strategy to avoid buying seeds that are not quality-verified. Yet, farmer knowledge about observable quality markers is low at baseline.

Second, the information treatment caused buyers to have better ability to detect seed quality. Treated buyers had greater knowledge of observable quality markers and were more likely to successfully use detection techniques in subsequent planting seasons. Treated buyers frequently reported that the information that was provided affected their purchasing decisions—both which seed packets to buy and which seed seller to buy from. Consistent with these accounts, we confirm that buyers in treated areas were more likely than buyers in untreated areas to leave the local market to purchase seeds. Examining agricultural outcomes, we find that treated farmers experienced about 5% higher maize yields overall, suggesting that treated farmers obtained higher quality seeds and this led to increased productivity. More remote farmers benefited especially (consistent with baseline seed quality being lower in such areas) as did more educated farmers (consistent with their demonstrating greater retention of the information that was disseminated).

Third, the intervention caused sellers to exit the market for hybrid maize seed in the subsequent two planting seasons, as revealed by data from secret shoppers that tracked seller responses for a little over one year after treatment. Among sellers that remained in the markets, we do not detect

[1] Among packets sampled by secret shoppers, 42% of them lack one or more quality marker.

treatment effects on seed quality offered to uninformed buyers.

These patterns can be explained by a model in which farmers who are informed can conduct a sequential search for high quality seeds–possibly leaving the local market to do so–and their increasing numbers hurt the profits of local firms. If the per-unit cost to the firm of upgrading quality from low quality to high quality is high enough relative to other firm costs, then firms cannot be induced to upgrade quality with improved consumer information. Instead, firms will prefer not to enter the market. Consistent with the model's predictions, we find that treatment caused approximately one in six sellers on net to exit the market, and we do not detect effects on prices or price dispersion.

Taken together, the results document new stylized facts and provide important evidence relevant for boosting yields of a staple crop. There is much work on barriers to adopting agricultural technologies that can enhance productivity. Jack (2013) surveys this literature, which spans studies on many types of constraints—including credit and liquidity constraints, risk aversion, lack of knowledge on costs and benefits, and heterogeneous costs, among other explanations for under-adoption—and has motivated policy responses to address market failures in the adoption of agricultural inputs (Holden, 2019). Some recent work has explored the role of low input quality as a barrier to improving agricultural productivity. Quantitative research (Ashour et al., 2019; Bold et al., 2017) and news reports (Muchiri, 2019; Okinda, 2019) suggest that the quality of agricultural inputs–such as seeds, fertilizer, and pesticides–is often low in rural African markets. For example, Bold et al. (2017) study retail quality maize seeds and fertilizer from rural Ugandan markets. They find that switching to wholesale quality of the same products causes a 40% increase in maize yields. These reports suggest that low and high-quality products often appear similar along observables (e.g. price and package characteristics), which is consistent with accounts of widespread counterfeiting (Kenya Association of Manufacturers, 2012). Our paper contributes to this literature by evaluating an intervention to improve consumer information as a way to boost usage of high quality inputs and improve agricultural productivity. The findings are directly relevant to policymakers working to promote food security. This paper also contributes important descriptive data on the quality of agricultural inputs (Bold et al., 2017; Gharib et al., 2021; Kenya Association of Manufacturers, 2012; Ashour et al., 2019; Michelson et al., 2021)), particularly on the relationship between observables and seed quality.

The paper also contributes to work related to consumer mistakes, consumer learning, and quality provision. There is a deep theoretical literature on market dynamics when agents have imperfect information (e.g. Akerlof, 1970; Wolinsky, 1983; Shapiro, 1983). However, relatively few empirical studies have tried to test theoretical predictions in a real-world setting (Bai, 2021; Jin and Leslie, 2009; Bjorkman-Nyqvist et al., 2012). We do so in a high-stakes and policy-relevant context, using an experimental evaluation closely tied to a common government policy (certification of minimum quality standards) intended to overcome a market failure arising from a difficult learning environment for consumers. This paper also relates to work on consumer misinformation and costly consumer mistakes (Bronnenberg et al., 2015; Grubb and Osborne, 2015; Handel and Schwartzstein, 2018). We document that many consumers overlook easy-to-use strategies to obtain higher quality products, behavior that is inconsistent with a full-information model with rational consumers.

Lastly, we contribute to work on regulation, monitoring, and enforcement. The information treatment under study is closely tied to processes for mandatory seed certification. Several papers examine how consumer information or monitoring by consumers can complement or replace direct enforcement efforts. Different from previous work (Gonzalez-Lira and Mobarak, 2021; Annan, 2021; Naritomi, 2019), we find evidence in our setting that consumers may be limited in their effectiveness in disciplining sellers, and we discuss the conditions from this setting that likely contributes to this finding.

2 Background

Maize (often called corn in North America) is a staple crop in many areas of the world. In Eastern and Southern Africa, maize is estimated to account for 22% of all calories consumed and is an important crop for subsistence farmers in the region (Shiferaw et al., 2011). A central challenge for policymakers concerned about food security is how to boost agricultural productivity in lower-income countries, where maize yields are a small fraction of the average in OECD countries (OECD/FAO, 2021). Increasing access to improved agricultural inputs such as high yielding seed varieties and fertilizers have been one focus among policymakers in recent decade (Evenson and Gollin, 2003).

However, recently scholarly work and local media coverage has suggested that retail maize seeds in East Africa and other agricultural inputs are variable and often sub-standard (Ashour et al., 2019; Bold et al., 2017; Muchiri, 2019; Okinda, 2019). In Kenya, the regulation of agricultural inputs falls under the responsibility of the Kenya Plant Health Inspectorate Service (KEPHIS). KEPHIS was established in 2012 by the Seed and Plant Varieties Act, and as part of its responsibilities, it tests and certifies crop seeds for sale. After manufactur, all seeds must be certified by KEPHIS before sale, and all sellers must also register with KEPHIS. Importantly, seeds must test with over 90% germination rate to be certified, meaning that under ideal temperature and moisture conditions randomly sampled seeds must develop normally. Without passing required checks to obtain certification, seeds cannot legally be distributed to wholesalers and retailers for sale. Concerns about counterfeit and other uncertified seeds led to a new initiative in 2018 to mandate e-verification for certified seeds, similar to verification schemes implemented in other setting (e.g. agricultural inputs in Uganda as in Ashour (2015)). This requirement for all seeds sold in Kenya was layered on top of pre-existing requirements, such as a printed lot number and packaging or expiration date printed on each packet. The new e-verification scheme allowed for an additional method to obtain and verify printed information–namely receiving key information about the seeds via SMS, which cannot be physically tampered with. Each packet is assigned a unique secret code that is revealed upon scratching off the sticker, and the code is available for one time use only. Subsequent uses would result in a message indicating that the code is valid but also alerting the user that it had been used before.[2]

Nevertheless, concerns about seed quality remain. As KEPHIS Managing Director Esther Kimani said, tying this issue to food security: "These fake seed sellers have...been the cause of food shortages that make Kenya spend billions of shillings on imports annually." Counterfeits may not be the only possible cause of low quality seeds – other reasons may include poor storage or selling seeds after expiration or too long after testing to ensure high quality. This is consistent with academic research on maize seed performance after storage under less-than-ideal conditions (Ghassemi-Golezani and Mamnabi, 2019). Testing for each seed lot that is manufactured is subject to testing requirements at KEPHIS. However, there are points within the supply chain between the seed manufacturer and the small retail outlets that we study, which allows for lower-quality seed packets to possibly enter the supply chain. First, the seed manufacturer relies on a network of agents and sub-agents to distribute the seeds, ending with the smallest retail shops and the farmers they sell to. Intermediaries and retail shops may store seed packets improperly, sell expired seed, or source seeds from unauthorized distributors, allowing for counterfeits or other non-quality-verified products to enter the supply chain. While the regulator tests seeds at the certification step that takes place after manufacture but before distribution, the regulator has limited capacity to monitor seed quality at the end of the distribution chain at which farmers purchase seeds.

Our study area is in Western Kenya in four counties - Bungoma, Busia, Kakamega and Transnzoia Counties. According to data from the 2014 Kenya DHS survey, 47% of households in the region experience food insecurity. In 2019, 74.5% of households in this area participated in farming activities, 92.4% of which farmed maize (Kenya National Bureau of Statistics, 2019).

3 A Simple Model With Informed and Uninformed Consumers

To structure our thinking about consumers and seller behavior, we consider a simple one-period model. In this model local sellers set price and quality levels in the presence of both informed and uninformed consumers who also have the option of purchasing from a seller outside the local market.

3.1 Demand side

Assume there are N consumers, all of whom seek exactly 1 unit of the good. The good comes in one of two quality levels: high or low. A consumer derives utility equal to a payoff associated with either high or lwo quality seed minus the price paid. For simplicity, we normalize this utility to $1-p$ if the consumer obtains a high quality good, and $-p$ if they obtain low quality. Let θ be the proportion of consumers who are informed. We treat θ as exogenous; one could think of the share of informed consumers as coming from a person-specific cost of acquiring information, which leads some to acquire this information and others to not acquire it.

Assume there is either 0 or 1 firm locally.[3] Consumers can buy from the local firm (if there is

[2] We thank staff at KEPHIS and mPedigree for helpful conversations about recent changes in seed regulations.

[3] Here we assume only up to one seller, which is clearly a simplification. However, the median market in our sample

one) at price p, or alternatively, they can buy from a firm outside the local market, paying search cost S and price P_0 for one unit of the good. We assume that the local market is small enough that P_0 would be negligibly affected by choices of local market participants. An informed consumer, upon taking the outside option, will always get high quality. An uninformed consumer believes they can get high quality with probability $\bar{q}$ in the outside option.

Informed consumers can observe quality markers and upon visiting the local shop will know the quality of the product (either high or low) with certainty. Uninformed consumers, on the other hand, cannot observe quality. They have beliefs about the average quality in the local market ($\hat{q}$). Note that we allow $\hat{q}$ to match the average quality offered by the seller, but we do not require that to be the case.Evidence suggest that buyers of agricultural inputs often do not hold accurate beliefs about product quality and that learning about quality through experience can be very difficult (Ashour et al. (2019); Bold et al. (2017); Hoel et al. (2022); Michelson et al. (2021); Michelson et al. (2023)). Learning through repeated experience may especially be hindered when the difference in quality between low and high quality is not very large. Thus we think of this as a model of short-run buyer and seller responses when beliefs about average quality change minimally.

We focus on the equilibrium in which the uninformed buy locally, and the informed buy locally *only* if they receive a high quality good. Since the informed expect to get high quality from the outside option at cost $P_0 + S$, and the uninformed expect to get quality $\bar{q}$ in expectation at cost $P_0 + S$, so the following must both be true in this equilibrium:

$$
\begin{aligned}
1 - p \geq 1 - P_0 - S \quad &\implies \quad p \leq P_0 + S \\
\hat{q} - p \geq \bar{q} - P_0 - S \quad &\implies \quad p \leq \hat{q} - \bar{q} + P_0 + S
\end{aligned}
$$

We might default to thinking of the second expression as binding, in which buyers believe they get better expected quality outside the local market (i.e. $\hat{q} < \bar{q}$). (That is, if the second expression holds it would imply that the first expression also holds.) Evidence from Gharib et al. (2021) would support this assumption, finding that farmers are willing to pay a premium for a seed packet directly from the seed company, which normally can be obtained outside the local market in town. Either way, in this equilibrium, the seller can raise prices up to $p^* = \min\{P_0 + S, \hat{q} - \bar{q} + P_0 + S\}$ without losing any of the uninformed consumers, or any informed consumers who get high quality. They lose informed customers who are offered low quality, who opt for the outside option.

3.2 Supply side

Let x be the number of units that the firm sells. Let $q \in [0, 1]$ be the quality mix chosen by the seller. That is, proportion q of the seller's stock is high quality, while proportion $1 - q$ is low quality. For each potential customer, the seller randomizes the quality of the product according to the choice of q, and they make a take-it-or-leave-it offer.

has only one maize seed seller. Evidence from some recent research has also suggested that sellers in similar rural settings in Kenya may collude and behave similarly to a local monopolist (Bergquist and Dinerstein, 2020)

The seller faces variable costs $(c + d\frac{q}{1-\theta+q\theta})x$, where c is the (assumed constant) per-unit cost for a unit of low-quality good, and d is the cost of upgrading a unit's quality from low quality to high quality. $\frac{q}{1-\theta+q\theta}$ is the average quality among units that are sold if the seller chooses quality mix q, but informed buyers refuse if offered low quality (which happens $q\theta$ of the time).

Let F be the firm's fixed cost.

Thus, the firm's problem is as follows, where we write x as a function of the firm's choice of price p and quality mix q:

$$Max_{p,q}(p - c - d\frac{q}{1-\theta+q\theta}) * x(p,q) - F$$

The firm's solution in the equilibrium of interest is as follows (see Appendix B for more details):

$$p^* = \min\{P_0 + S, \hat{q} - \bar{q} + P_0 + S$$
$$q^* = \begin{cases} 1 & \text{if } (p-c)\theta - d > 0 \\ 0 & \text{if } (p-c)\theta - d < 0 \end{cases}$$

In other words, $q^* = 1$ when $\theta > \frac{d}{p-c}$. When the fraction of consumers that are informed is large enough, the firm is incentivized to provide high quality. Higher upgrade cost and a smaller profit margin will tend to push the threshold θ must cross higher, making it more difficult to attain a level of consumer information that will induce sellers to upgrade quality.

Meanwhile, firm profit is always at least weakly decreasing in the fraction of consumers that are informed:

$$\frac{\partial \pi^*}{\partial \theta} = -(p-c)(1-q^*)N \leq 0$$

Anticipating this, the firm will not enter the market if expected profits π^* falls below a certain value representing its opportunity cost:

$$(p-c)(1-\theta)N - F < \pi_0$$

Re-writing this, we see that the firm will not enter the market if:

$$\theta > 1 - \frac{F + \pi_0}{(p-c)N}$$

As θ rises, does the firm improve quality or exit? Putting together the inequalities above, we can show that the firm will improve quality rather than quit as θ rises if and only if:

$$d < (p-c) - \frac{F + \pi_0}{N}$$

3.3 Summary of model predictions

1. Buyers that become informed obtain higher quality goods. By the assumptions made, informed buyers can observe quality and are therefore able to search until they find high quality, leaving the local market if necessary to find another seller.
2. Having more informed buyers does not affect market prices ($\frac{\partial p}{\partial \theta} = 0$). Sellers have local market power, and prices are bounded from above by uninformed consumers' value of buying from the outside option.
3. Having more informed consumers has a negative effect on local sellers' expected profit ($\frac{\partial \pi}{\partial \theta} \leq 0$). More informed consumers reduce the local seller's market share
4. Uninformed consumers may also benefit from having more informed consumers in the local market. They benefit if the seller opts to upgrade quality rather than quit as the number of informed consumers increases. This happens if upgrade cost isn't "too high" (relative to a function of the firm's profit margin and opportunity cost of selling seeds).

4 Study Design

4.1 Sample

We randomly sampled rural markets that satisfy the following conditions: (1) it must have fewer than 100 shops, (2) it must be more than 2km from a market that has more than 100 shops, (3) it must have at baseline at least one seller of maize seeds. We consider markets that satisfy these conditions eligible to be sampled. We sample markets from four counties: Bungoma, Busia, Kakamega, and Transnzoia using a 2-stage sampling strategy–first randomly selecting sublocations (a list of which we obtained from the county commissioner offices), next tabulating all eligible markets in selected sublocations, and lastly randomly selecting one market (or two in randomly selected sublocations in Transnzoia, due to logistical reasons). This sampling strategy helps minimize clusters of markets that are very close to one another, which would possibly risk spillovers between treatment and control sites, and it helps the sample of markets cover both less remote and more remote areas of each county. The main study sample in treatment and control groups consists of 302 markets in 282 sublocations.

At each market, all seed sellers and eight randomly selected maize-farming households within 1km from the market center were sampled to be surveyed. The sampled farmers are overwhelmingly small-holder farmers, with the median farmer planting on 1.5 acres. The team did not conduct a census of all households within the sampling frame. Instead, enumerators used systematic random sampling, starting at a randomized location and walking along roads and footpaths according to a set pattern. Every nth household was selected to be surveyed, with the skip interval set at one-sixteenth of the estimated number of households residing within one kilometer from the village center (as reported by the village elder).

In addition to markets in the treatment and control groups, 104 pure control markets were also selected using the same methodology, except that we do not observe the number of seed sellers in Feb-Mar 2020. We instead use the number of seed sellers in March 2021. For these sites we collected only data from market audit activities during the 2021 main planting season. Due to funding limitations, we did not conduct seller surveys or household surveys in pure control markets.

4.2 Randomized experiment

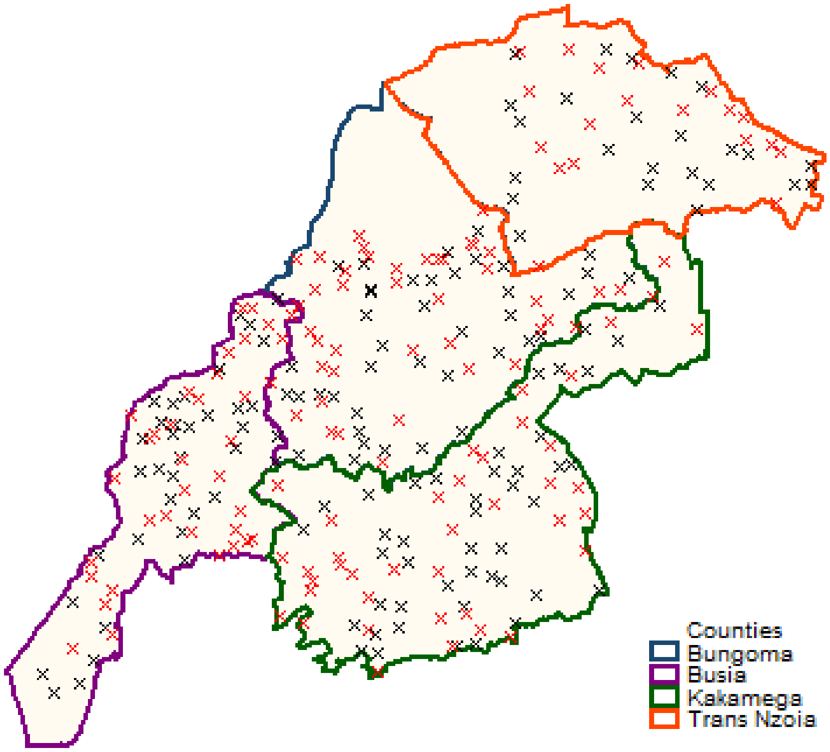
Sublocations were randomized into the following treatment groups, and we describe the treatment arms in greater detail below.

Treatment #1	: 68 sublocations
Treatment #1 & Treatment #2	: 68 sublocations
Control	: 146 sublocations
Pure Control	: 104 sublocations

Figure 1 shows the sampled markets within the four counties, with treated markets in red. The main information treatment (Treatment #1) provides information to help consumers distinguish between low and high-quality products. Through pilot activities and discussion with project partners, we identified a set of quality markers that are required for all certified seeds in Kenya and have variation in rural markets. These include the lot number and a sticker containing an SMS code that allows farmers to verify authenticity (see Appendix A). In response to widespread reports of deceptive counterfeiting, starting in 2018 KEPHIS required all certified packets of seed to carry one of these scratch-off codes. Using SMS verification allows farmers to detect two types of lower-quality seeds: uncertified seeds and old seeds that were unsold from previous seasons. Hybrid seeds are typically purchased in closed packets produced by the manufacturer. Most of the quality markers associated with Treatment #1 can be observed for all brands of seed prior to purchase without opening the packet, and in practice these are the ones for which we observed meaningful levels of variation across sellers. Treatment #2 encourages consumers to report sub-standard seeds when they encounter them. Consumers were (1) encouraged to ask for a receipt and keep the packaging for documentation, (2) told how to report the incident anonymously to the Anti-Counterfeit Authority, a corporate representative, an agricultural officer, KEPHIS, or the local chief (who were informed by the team how to escalate reports), and (3) encouraged to discuss seed quality among friends, family, and neighbors. They also were told of documented instances where complaints led a company to give compensation or led to legal action.

Both treatments were administered in the month immediately before the main planting season in 2020, which occurs around the start of rains in March. The team carried out the campaign by working with the local assistant chief to deliver flyers to and speak to locally influential residents at gatherings, including village elder meetings, local farmer group meetings, and barazas and chamas (local community meetings). The team also spoke with individual farming households, going door-to-door to deliver flyers and speak about the quality markers. Survey respondents in treated communities were also given the information treatment; this was done immediately after the baseline survey was concluded.

Figure 1: Study Area



The study area in Western Kenya includes Bungoma, Busia, Kakamega, and Transnzoia Counties. Red X's represent treatment sites, where community-wide training to help consumers detect quality-verified seed was carried out. Black X's represent control sites. In total 302 market areas are included in the sample.

All treated sites received Treatment #1, which was designed to increase the probability of detecting non-quality-verified seeds, corresponding to an increase of θ in the model. The model predictions from the previous section guide our thinking on the expected effects the information treatment. As illustrated in the model, this can directly affect purchasing decisions (e.g. refusing a packet or switching sellers) and lead to the adoption of higher quality seeds. Treatment may also affect sellers' decisions (e.g. selling more high-quality seeds or adjusting entry and exit decisions) and lead to a greater prevalence of high-quality seed in the market for even uninformed buyers. This may happen through economic channels as buyers refuse to buy when observing a poor signal (as in Wolinsky 1983).

Appendix A shows the (English version of) flyers which summarize the information that was communicated at treated market areas.[4] It also shows typical packets of seeds and screenshots from using the e-verification system. (Those who received flyers were encouraged to pass extras along to family, neighbors, and friends.) Focus group discussions prior to project launch and treatment meetings suggest that farmers perceived the provided information to be important for their livelihoods and that the information is accessible, particularly with the help of family, friends, and neighbors. We note that one of the quality markers–the KEPHIS sticker–is more technically challenging to use

[4] In practice, nearly all distributed flyers were the Kiswahili version.

properly. While the application of SMS codes to quality verification may be new to many farmers, we note that the technology is often familiar from its widespread use for pre-paid mobile phone service, which may explain the ready acceptance of this aspect of the treatment.

Treatment intensity was not uniform across treated sites, with two-thirds of treated markets receiving a more intensive multi-day treatment (two or three days of information dissemination). We implemented this design including multiday treatments out of concern that treatment effects may only be seen for a large enough change in θ. For example, consider the model from section 3 in which some markets begin with sellers offering $q = 0$. In such cases, only a large-enough change in θ that causes a threshold to be crossed will induce a change in the quality provided by sellers or induce sellers to exit the market. In practice, though, the data shows little evidence of such threshold effects from the variation in treatment intensity. Therefore section 5 below will primarily focus on simple treatment-control comparisons in order to maximize statistical power.

Treated market areas received on average 173.5 flyers, which we estimate as being approximately 19.7% of local seed customers (see Table D1).[5] On average in treated markets the team trained 81.7 adults face-to-face, or about 9.1% of local seed customers. Considering village elders' estimates for the local population, about 49% of the local (within 1km) adult population received a flyer, and 22% were treated directly face-to-face.

4.3 Data Collection and Timeline

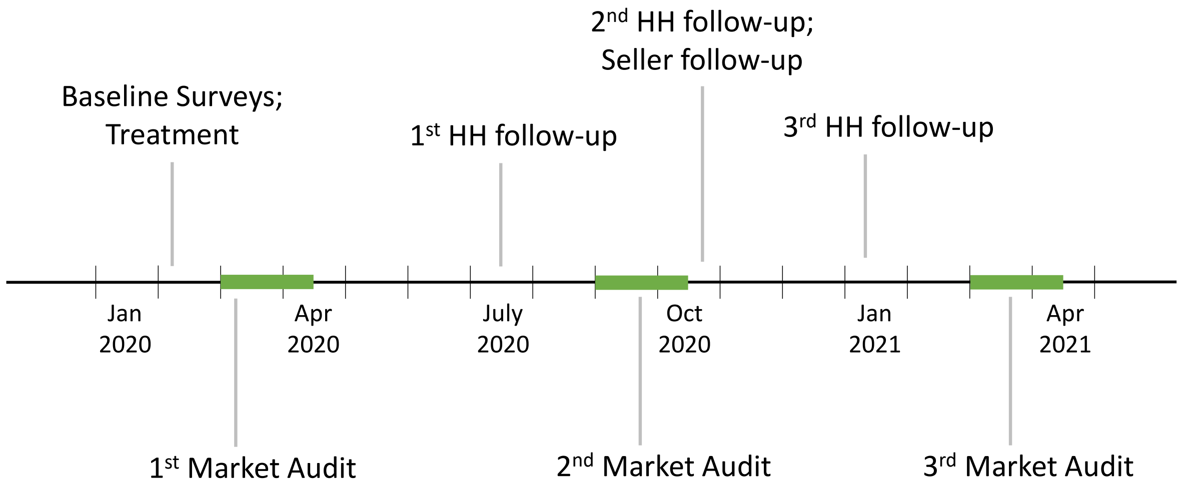
Project activities proceeded as illustrated in Figure 2. In February 2020 into early March 2020, baseline surveying and treatment were carried out. After a site was initially sampled and the team made contact with village elders, baseline household and sellers surveys were carried out. The information treatment was normally delivered later in the day during the first day of surveying, with participants convening in the mid-afternoon in a public location as arranged by the village elder. Market areas that are assigned to receive additional visits for information dissemination would then have a team member return for additional activities on subsequent days.

Over the following one year the team conducted three rounds of household follow-up surveys via phone to document household seed choices and agricultural outcomes. Households self-reported harvested amounts for each maize variety they planted, which were used to compute maize yields for each household. Survey times were spread out strategically around key dates in the usual crop timeline in order to minimize issues with imperfect recall of key variables (e.g. seed purchases and harvested amounts) for both the main season (with planting starting around March 2020) and the short season (with planting starting around September 2020).

To measure seed quality and possible adjustments by sellers, we conducted three rounds of market audit activities. In March 2020, August to September 2020, and March to April 2021, covert shoppers (enumerators on the research team) posed as farmers to visit markets and view up to two packets and purchase up to one packet. They used realistic scripts to purchase "seeds

[5] We use several pieces of data to estimate this figure, including: number of sellers in a market, number of customers per seller, and percentage of local customers who shop locally for seeds.

Figure 2: Timeline



The project activities included: (1) baseline surveys and information treatment conducted starting in February 2020; (2) three market audit activities conducted starting in March 2020, September 2020, and March 2021; (3) a seller follow-up survey to capture prices, quantities, and profits of sellers; (4) 3 rounds of household follow-up surveys, designed to capture key information about seed purchases and agricultural outcomes in both the main growing season and the short growing season.

that are popular" or "seeds that are cheap", without naming the exact brand or variety that is desired, leaving the seller to decide which seeds to offer. In the local context, requesting seeds in this way without specifying the exact variety is perceived as normal. When outside of the market area, the secret shopper was instructed to document visible quality markers as well as data about the sellers in the market and the shopping experience. Conversations with field staff lead us to expect that secret shoppers were not recognized or treated differently from other shoppers during these relatively quick transactions.

All seed packets were repackaged into plain paper bags, labeled, and sent to KEPHIS facilities in Nakuru for purity and germination testing, to provide objective measures of two aspects of seed quality. First, purity tests measure the percentage of material (by weight) in the packet that are whole seeds. Second, germination tests measure the percentage of whole seeds that become normally emerging plants under ideal temperature and moisture conditions. While these tests provide an objective test of seed quality, with KEPHIS staff blinded to most seed characteristics that are not visible on the seed itself, it does not capture all relevant aspects of seed quality that a farmer may care about. These lab tests do not confirm the seed variety through DNA testing, provide information about seed performance when conditions are less-than-ideal, nor do they capture aspects of seed quality that affect yields beyond the germination rate. As demonstrated by Ghassemi-Golezani & Mamnabi (2019), it is expected that older or poorly-stored maize seeds will have lower yields due to deterioration of seed quality beyond the effects on germination rate alone.

The team documented the following markers during the market audit: (1) presence of valid SMS verification code, (2) presence of lot number, (3) testing date within 1 year of purchase, (4) damaged packaging (which reports have linked to tampering). A packet can lack a valid SMS verification code several ways. First, a KEPHIS sticker could be absent. Second, the code could be invalid. Third, the code could be valid but already used prior to purchase. Lastly, the code could be successfully applied for the first time, but the registered information may not match the information on the packet. Similarly, a packet could lack a valid expiration date either because it lacks any date whatsoever, or because it has a date but the date indicates that the seeds are expired.

During follow-up surveys, enumerators were not made aware of the treatment status for the sites and respondents that were being followed up on. Lab testers were not made aware of the seed sources; all seed samples were repackaged into plain paper bags labeled with unique alpha-numerical identifiers used internally by the research team and do not reveal any seed characteristics.

One concern with the survey data may be that measurements on agricultural outcomes are self-reported and could suffer from inaccurate reporting. To guard against these concerns, we used self-reported yields to inform our sample size, anticipating possible noise from mismeasurement.[6] Acreage, kilograms of maize harvested, and kilograms of seed purchased are also checked through repetitive questions in different parts of the survey and cross-validated with each other. For example, the total amount of land dedicated to maize should be consistent with the sum of land dedicated to each variety of maize. The surveys also phrase questions and enumerators use probing in ways that help respondents recall the answers we are seeking. For instance, we ask "how many 90kg bags of maize did you harvest during the long growing season?" as a question respondents often readily recall.

4.4 Baseline Balance, Spillovers, and Other Threats To Identification

Table D2 shows baseline balance for market areas in our sample. Households in treated sublocations had households heads that were slightly less likely to have completed primary or secondary school, and they had slightly smaller plots. However, when we jointly test for differences between treatment and control on all of these measures, we cannot reject the null hypothesis that all characteristics are the same in both treatment and control groups (p=0.44). To address the concern that imbalance in

[6] We used survey data from the Tegemeo Institute.

these variables could skew the experimental results, the results below showing treatment effects will control for these baseline characteristics, together with county and planting-season fixed effects.

In the Appendix C, we explore the possibility that the experimental design may have been contaminated in one of two ways. First, did baseline activities influence sellers by altering their beliefs about the likelihood their products will be scrutinized? Second, did information spillover from treated sites into neighboring market areas and affect buyer and seller knowledge and behavior there? We do not find evidence consistent with these two hypotheses and so in section 5 below we primarily focus on comparisons between treatment and control sites.

5 Results

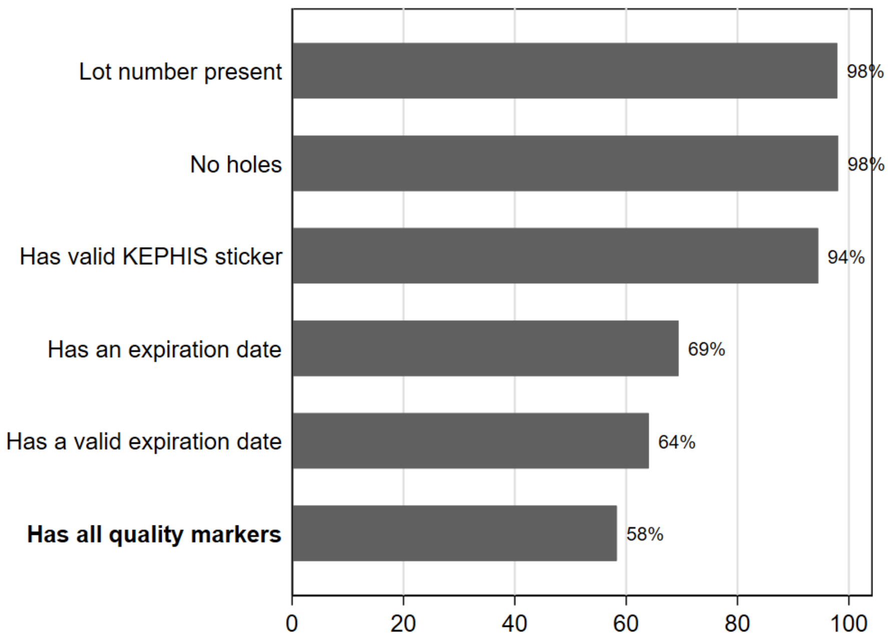
In this section we present the main results as follows. In subsection 5.1, we confirm that observable markers correlate with lab-tested germination rates. This suggests that farmers can indeed use simple purchasing strategies based on observables to help them acquire better quality seeds and in theory achieve agricultural gains. In subsection 5.2, we describe treatment effects on buyer knowledge and usage of quality markers, and seed purchases. The findings suggest that informed farmers are able to adjust their purchasing behavior, consistent with channels in the model. In subsection 5.3 we examine agricultural outcomes. We find that treatment improved maize yields, particularly among more remote farmers (where quality markers were associated with greater gains in germination rate), and among more educated farmers (who appeared to retain the information better). This finding corresponds to model prediction #1. The results also show that prices paid by locally-buying farmers are not affected, corresponding to model prediction #2. In subsection 5.4, we examine effects on sellers' decision to enter or exit the hybrid seeds market. We document a sizable effect of treatment on seller exit, consistent with the negative effects of informed buyers on profit from model prediction #3. Lastly in subsection 5.5, we examine effects of information on prices and quality for uninformed buyers in the market. We do not find effects on seed quality for the uninformed, which correspond to model prediction #4 in a world where upgrade costs are high relative to other firm costs.

5.1 Summary statistics: quality markers and their correlates

We first describe summary statistics for the data on quality markers that was collected. These descriptive statistics provide new insight into the prevalence of various observable quality markers and their correlates. Positive correlation between observable markers and objective lab-tested quality measures suggest that KEPHIS regulations were implemented with some success and that farmers stand to benefit if they use these markers to decide which seed packets to purchase.

Figure 3 presents summary statistics for each quality marker. Overall, 58% of packets that our team observed had all quality markers, while 42% were confirmed to be missing one or more. Most of the missing quality markers were due to not having an expiration date, or not having a valid expiration date on the packet; missing or invalid KEPHIS stickers and the other observable markers also contributed somewhat to missing quality markers.

Figure 3: Quality marker frequency

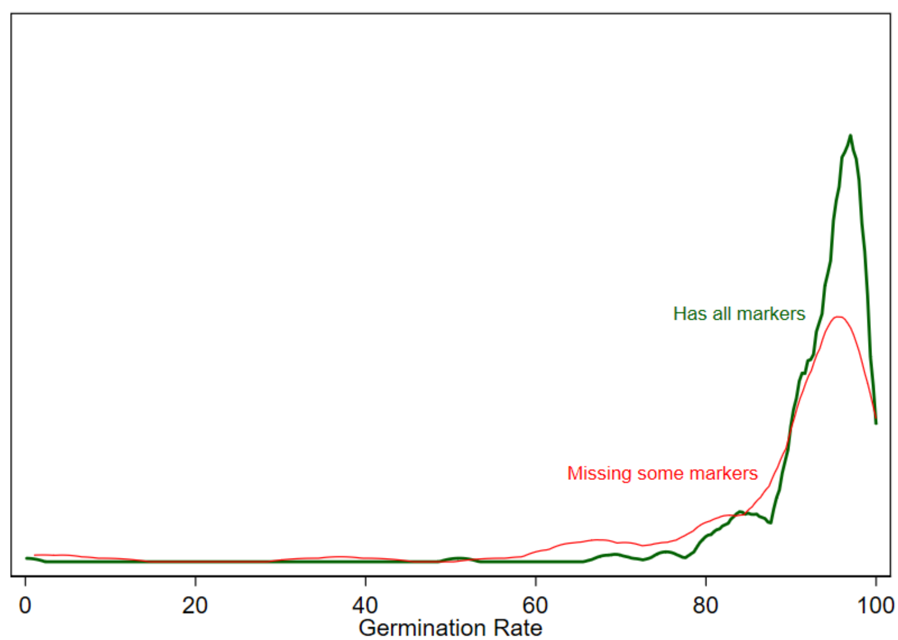


This figure shows the percentage of seed packets that feature each quality marker, calculated using data from market audit surveys which were submitted by secret shoppers on our research team. The sample includes 1508 seed packets. This includes observations from all three planting seasons in which market audit activities were conducted. It also includes both packets that were purchased as well as packets that were closely observed but not purchased. In the latter case, enumerators were not able to scratch the KEPHIS sticker and fully complete the e-verification steps. We categorize those packets as having a valid KEPHIS sticker so long as all other quality markers are present, and the sticker was present with all visible features of the sticker appearing to be valid. Overall, we find that 62% of packets had all quality markers present.

We next ask if observable quality markers correlate with aspects of seed quality measured in lab tests. Figure 4 shows that packets with all quality markers tend to have higher germination rates than packets that are missing one or more quality marker. Table 1 shows that observable quality markers are strongly correlated with germination rate but not purity. Table D3 and Table D4 break down the analyses by quality marker. This shows that the the relationship between quality markers and germination rate is driven primarily by two of the quality markers: (1) presence of a KEPHIS sticker, and (2) presence of a valid expiration date. Other aspects of seed quality are likely positively correlated with germination rate (Ghassemi-Golezani and Mamnabi, 2019), and so we view this as a lower bound for the yield increase that is possible if non-quality-certified seeds could be eliminated from the marketplace.

We view the observed 5.2% overall difference in lab tested germination rates as a lower bound for the changes in yield that could be achieved should farmers exert perfect discretion in choosing seed. It does not account for other characteristics of seeds that affect yield, which is expected to correlated positively with germination rate and whose effects on yield may far exceed the differences in germination rate (Ghassemi-Golezani and Mamnabi, 2019). This relatively small magnitude for differences in germination rate (which is more easily observable to farmers than other seed characteristics), could help explain the persistence of the baseline market equilibrium. If low-quality seeds were even more inferior, farmers would be able to learn more quickly, which works to unravel the pooling equilibrium that we observe in which both lower and higher quality seeds are sold by sellers at the same price.

Figure 4: Distribution of Germination Rates



The sample includes 467 seed packets that were purchased by secret shoppers from a control site and tested in a lab by KEPHIS staff. The sample includes packets purchased during the long growing season in 2020, the short growing season in 2020, and the long growing season in 2021.

The relationship between quality markers and lab-tested germination rate differs in more remote versus less remote markets. Table 2 shows this relationship for market areas that are farther than the median distance from the county capital, as opposed to markets that are closer than the median distance. In more remote market areas, germination rates are lower overall and farther from national standards for minimum germination rate. Also, in these areas the quality markers are also more informative to farmers in the sense that having all quality markers is associated with a greater gain in germination rate in more remote markets than in less remote markets.

These correlations between observable markers and lab-tested germination rates do not capture all relevant aspects of seed quality. However, we expect germination rates to correlate with yields, and the tests provide high-quality objective evidence that farmers – particularly in more remote areas – stand to gain in agricultural output if they can successfully use these observables to choose

Table 1: Quality markers, seed purity, and germination rate

	Purity	Germination Rate
Has all quality markers	0.0	5.2***
	(0.0)	(1.1)
Constant	99.8***	88.1***
	(0.0)	(1.0)
Observations	467	467

The sample includes 467 seed packets that were purchased by secret shoppers from a control site and tested in a lab by KEPHIS staff. The sample includes packets purchased during the long growing season in 2020, the short growing season in 2020, and the long growing season in 2021. The dependent variable in column 1 is the estimated percentage of material that are whole maize seeds (in percentage points). The dependent variable in column 2 is percentage of seeds that germinated in the lab (in percentage points). Table D3 and Table D4 show similar results for the complete list of quality markers.

higher quality seeds. Lab tests are not subject to self-reporting biases and testing staff are blinded to the seed source, variety, and any characteristics of the packaging. This also reflects a sort of limited success in the KEPHIS certification process – implementation failures or widespread counterfeited marks have not caused observables to be uninformative about seed quality.

5.2 Effects on household knowledge and seed purchases

The following sections show treatment effects as estimated by the following specification:

$$y_{isc} = \beta_0 + \beta_1 \text{Treated}_i + X_{isc}\Gamma + \delta_c + \varepsilon isc$$

Here y_{isc} is the value of an outcome for household i, during season s, in county c; $Treated$ is a dummy that equals 1 if the sublocation in which the household resides was assigned to the treatment group; X_{isc} is a vector of controls, which as described in the pre-analysis plan includes the household head's gender and age, and the value of the dependent variable at baseline. In the specifications shown below, we control for county fixed effects, planting-season fixed effects, and baseline characteristics shown in Table D2. Since for example sublocations are of varying sizes, especially in different counties, observations are weighted inversely to the probability that their associated market is selected to be included in the sample so as to produce results that are representative of markets in the study area. Appendix D shows results of robustness checks, which show the main takeaways are qualitatively unchanged when using alternative specifications for estimation. Appendix D also shows results examining heterogeneous effects along pre-specified dimensions, focusing on differential impacts by remoteness of the market, gender of household head, education of household head, and plot size.

We examine treatment effects on several measures of household knowledge, as shown in Table 3. Treatment substantially increased knowledge about visible quality markers, being associated with increased recall of specific elements by 35% to 120%. We note that household-level effects appear

Table 2: Remote markets have more to gain

	Germination Rate
Has All Quality Markers * Not Remote	3.3**
	(1.5)
Has All Quality Markers * Remote	7.5***
	(1.8)
Remote	-5.5***
	(1.9)
Constant	90.5***
	(1.3)
Observations	467

The sample includes 467 seed packets that were purchased by secret shoppers and tested in a lab by KEPHIS staff, and were collected from control markets. The dependent variable is the percentage of maize seeds that germinated in the lab (in percentage points). Remote markets are defined as markets with above-median distance to the county capital. "Has all QMs" equals 1 if a seed packet has all quality markers, and equals 0 otherwise.

to be heterogeneous. For households with more highly educated households heads with primary school or secondary schooling completed, knowledge (by all measures) both started at a higher level as well as increased substantially more due to treatment (Table D5).

We also examine the possibility that households at control sites near treated sites may have experienced gains in knowledge due to spillovers. The data suggest that any spillover effects that may have occurred are too small to be detectable. As shown in Appendix C, we do not find evidence of spillover effects within 2km, 4km, or 6km. Partly by design, due to the multi-stage sampling strategy, the sample does not have many markets that are clustered very closely together (e.g. within 2km of one another). This means that few markets are at closer distances where we might be concerned that spillovers could greatly affect estimates of treatment effects. It also means that the analysis does not have statistical power to detect spillovers across very short distances, where one might expect such spillovers to be strongest. Nevertheless, to the extent that undetected spillovers affect control sites in the region similarly but to a smaller degree than the treatment sites, one could think of the estimated treatment effects below as being lower bounds.

A priori, one concern that could limit the effectiveness of treatment is that many farmers could already possess the information we seek to disseminate. The results show this is not the case and they confirm that treatment was able to increase knowledge of quality markers, from a relatively low baseline. While possibly discouraging from the point of view of the regulator, it is encouraging for our research design that we find low awareness for the SMS-based verification system, suggesting that the intervention was well-timed to publicize information for impact on market outcomes. As our contacts at KEPHIS explained, they have yet to run a major campaign to inform buyers of the verification system, while the stickers themselves say little about their purpose (see Appendix A). As shown in Table 3, control households had little knowledge about quality markers–on average naming only 0.56 markers, and with less than 7% of farmers citing the seeds' expiration date–reflecting the low levels of informed consumers at baseline.

Given the timing of data collection, we think it is highly likely that the increases in knowledge represent lower bounds on the true increase in knowledge due to treatment that is relevant at the time of seed purchases. These data were collected in July 2020 at the earliest, and respondents did not know in advance when they would be called by the research team. While originally planned for immediately after the planting season ended in late April, field operations were disrupted by the onset of the COVID-19 pandemic, and so the first post-treatment measurements of knowledge come from a phone survey with the main respondent approximately five months after treatment.

Table 3: Household knowledge and usage of markers

Panel A: Household Knowledge

	(1) no. quality marks known	(2) e-verif known	(3) expiration known
treated	0.227***	0.118***	0.028**
	(0.040)	(0.014)	(0.011)
Observations	4456	4456	4456
Control Mean	0.56	0.13	0.07
Treatment Effect (%)	40.10	89.86	41.06

Panel B: Usage of markers

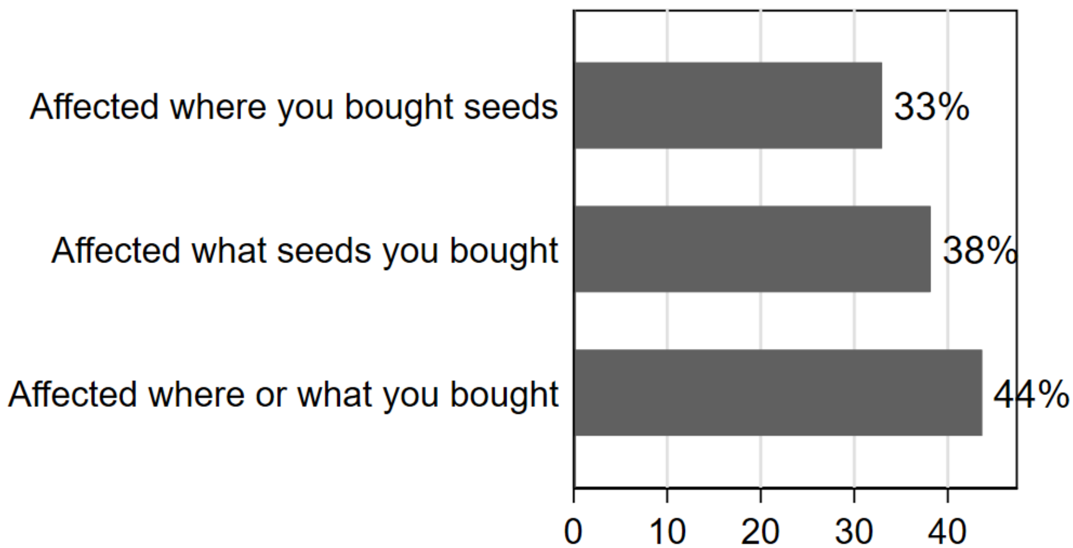
	(1) Able to verify	(2) Used e-verif successfully
treated	0.083***	0.092***
	(0.014)	(0.012)
Observations	4456	4456
Control Mean	0.14	0.07
Treatment Effect (%)	60.67	129.50

The table shows the effects of information treatment on household knowledge and household usage of markers. The sample includes observations from both the main planting season in 2020 and the short season in 2020, with knowledge measured approximately in late July 2020 and in November 2020. Standard errors are clustered by sublocation. In Panel A, the dependent variables are (1) the number of quality markers named by the respondent, (2) whether e-verification was named as a way to verify seed quality (=1 if yes), (3) whether expiration date was named (=1 if yes). In Panel B, the dependent variables are (1) whether the respondent reported being able to verify that seeds from that season were of high quality, and (2) whether e-verification was used successfully.

After the planting season, we asked treated households whether the treatment affected their seed purchase decisions. Respondents were asked the following questions: (1) "Did the information we provided about quality markers help?", (2) "Did it influence your decision of where to buy seeds? Please explain", (3) "Did it influence your decision of which seeds to buy? Please explain". Figure 5 summarizes their responses, in which 44% of respondents said that the information affected what seeds they purchased or where they purchased them.

Table 4 shows results examining whether treatment affected whether and where households bought hybrid maize seed. Columns 1 and 2 show that households shifted their source for seeds, with 4.9 percentage points fewer households sourcing seeds from the local market, and a similar amount obtaining seeds elsewhere. Gharib et al. (2021) provides complementary results that suggest farmers

Figure 5: Do treated households think the information helped?



The sample includes 742 respondents in treated market areas. These respondents were asked the following questions: (1) "Did the information we provided about quality markers help?", (2) "Did it influence your decision of where to buy seeds? Please explain", (3) "Did it influence your decision of which seeds to buy? Please explain". 44% of respondents indicated that the information that was provided affected either where or what seeds were purchased.

have higher willingness to pay for seed packets that have certain quality markers present.[7] These results are consistent with the mechanisms in the model, in which informed buyers that cannot obtain high quality seeds locally find it worthwhile to continue their search by buying outside the local market.

Overall, however, households were not more likely to buy hybrid seeds, as seen in column 3 of Table 4. It is possible that even if farmers updated expectations about the quality of seeds that they can expect to get (when using the quality markers), the shift is not near the threshold of pushing expected economic returns to be positive for many households. Another consideration is that additional market imperfections (such as credit constraints or incomplete insurance markets) may act as important barriers that limit movement on the adoption margin (Karlan et al., 2014).

[7] Gharib et al. (2021) estimates the effect of training (on two of the seven quality markers used in this paper) on willingness-to-pay for maize seed packets.

Table 4: Household Seed Choice

	(1) Bought hybrid at local market	(2) Bought hybrid elsewhere	(3) Bought hybrid
treated	-0.049**	0.073***	0.021
	(0.024)	(0.023)	(0.017)
Observations	4404	4456	4456
County FE	Yes	Yes	Yes
HH Baseline Controls	Yes	Yes	Yes
Control Mean	0.31	0.27	0.59
Treatment Effect (%)	-15.95	26.94	3.51

This table shows treatment effects on household seed purchase decisions. The sample includes households in both the main season in 2020 and the short season in 2020. The dependent variable in column 1 equals 1 if the respondent bought hybrid seeds at the local market in that planting season (and 0 otherwise). The dependent variable in column 2 equals 1 if the respondent bought hybrid seeds from a location other than the local market in that planting season (and 0 otherwise). The dependent variable in column 3 equals 1 if the respondent bought hybrid seeds from any source in that planting season. All standard errors are clustered by sublocation.

5.3 Effects on household agricultural outcomes

We next examine how the information campaign affected household agricultural outcomes. Total harvested amount and yield (kilograms harvested per acre) are the main outcomes we examine in this section. While inherently noisy and affected by other farmer inputs, these agricultural outcomes capture important elements of seed quality that are not measured in the lab tests that were conducted.

Table 5 shows that treatment caused an increase in reported harvested amount and yields. Kilograms of maize harvested per acre increased by about 6% due to treatment. To check the implications of these yield gains, let's assume that 42% of packets do not have a quality marker (average in control group for secret shoppers), and the average household experiences a 6% increase in yield. Assuming that all informed households reject non-verified packets in favor of quality-verified packets, as recommended in the training, then this implies that informed buyers who got a different quality level than in the counterfactual saw a 14% gain in yield, while the remaining buyers would have gotten high quality even in the counterfactual and had a gain of 0%.

For reference figures to help put these effect sizes in context, we can look to related research. Fabregas et al. (2019) does a meta-analysis of 7 digital agricultural extension interventions and find an average of 4% increase in output. Bold et al. (2017) finds a 13-18% increase in yield when switching from retail to wholesale quality seeds; they find a 28-38% increase in yield when switching from local to wholesale quality seeds.[8] Ghassemi-Golezani and Mamnabi (2019) used artificial aging to lower seed quality and obtained seeds that had 2% or 7% lower germination. Compared to these comparison groups, the higher quality seeds had 23.5% or 64.5% higher yield. Thus, the magnitude of our results are comparable to other agricultural informational interventions, and are comparable

[8] Even larger increases were seen when fertilizer quality was also increased.

with what might be expected (given evidence from poor quality seeds in Uganda, or evidence on the performance of old seeds from controlled trials), especially considering that farmers likely had imperfect usage when applying the training to seed purchases.

We do not see any effects on the prices that households paid (Table D7), nor do we see evidence of greater price dispersion among treated households (Table D8). This suggests that the estimated agricultural gains led to welfare gains for the households.

We have seen that treatment effects on lab-measured germination rates are weakly positive at best and appear insufficient to explain measured increases in yields. There are a few possibilities to reconcile these results. One possibility is that uninformed consumers in treated market areas (which secret shopped mimic) do not in fact obtain higher quality hybrid seeds as a result of treatment. This is perfectly consistent with informed consumers obtaining higher quality seeds, as is a feature of the toy model when firm upgrade costs are high. Another possibility is that improvements in germination rates are relatively small, and not detectable with our sample size. Research shows that other aspects of poor quality seed can affect yields in ways other than those mediated by germination rates; as a result, lower quality maize seeds can feature relatively small declines in germination rate and at the same time declines in yield that are many times larger (Ghassemi-Golezani and Mamnabi, 2019). Lastly, expired seeds do not germinate as well, while non-hybrid seed may germinate as well as hybrid seed but will have less yield. We cannot rule out the presence of well-germinating non-hybrid seed in non-certified seed packets; this can lead to minimal differences in germination rates while yields are substantially affected.

Treatment effects are concentrated in more remote areas, as illustrated in Figure D1, which shows results of four pre-specified analyses of treatment effect heterogeneity. In more remote areas, we note that baseline seed quality was worse while quality markers are associated with greater quality gains (Table 2). Households in more remote areas appear similar socioeconomically (e.g. education levels, housing quality, plot size), though we note that more remote farmers tend to have higher yields and lower population density. We also expect more remote areas to be less intensely monitored by regulators, which would potentially be a reason for seeing the lower baseline quality of seeds. Figure D1 also shows somewhat higher gains for households with a more educated household head. This is consistent with the observations that more educated respondents were better able to retain the information contained in the training, as seen in follow-up surveys that assessed their knowledge. The figure also shows somewhat larger gains for female-headed households, which is suggestive that resolving information frictions may help close the gender gap in agricultural productivity in LMICs (Diiro et al 2018; Wambua et al 2018). This finding also mirrors a result from Annan et al (2021) in which female customers in Ghana benefit more than male customers from an information treatment that helps them avoid being defrauded by mobile money agents. Findings are similar when dropping households who reported purchasing some seeds prior to the baseline survey, who might be hypothesized to respond less to treatment (Table D13), in markets with different levels of baseline competition amongst seed sellers (Table D14), when dropping respondents that had a close personal relation with their usual seed seller (Table D15), and when using an ex-post double machine learning approach to select control variables following Chernozhukov et al. (2018) (Table D16).

An important caveat to these results is that we cannot rule out changes in other complementary inputs (e.g. labor input, quality of fertilizer and pesticide, etc). We can however to check whether the information treatment affected usage of fertilizer, and we do not see effects (see Table D6).

Table 5: Household agricultural outcomes (All seasons)

	(1) Kgs harvested	(2) Yield
treated	57.75**	54.84**
	(29.05)	(27.39)
Observations	3807	2443
County FE	Yes	Yes
HH Baseline Controls	Yes	Yes
Control Mean	551.59	846.88
Treatment Effect (%)	10.47	6.48

This table shows treatment effects on household agricultural outcomes. Results include households in both the main season in 2020 and the short season in 2020. Yield is measured as kilograms harvested per acre. All standard errors are clustered by sublocation.

5.4 Effects on seller entry/exit

We next examine the effect of information treatment on seller entry and exit. The dependent variable in this analysis is the number of sellers present at the market. [9] As shown in Table 6, the number of sellers per market decreased by about 0.27 sellers in the SGS2020 and LGS2021 seasons due to treatment. This is about 17% percent decrease in the number of sellers.

The pattern of sellers exiting due to treatment is consistent with model prediction #3, where informed buyers cause sellers to expect increased costs and fewer customers, which drives down expected profits. The pattern of sellers exiting due to treatment is also confirmed in our separate dataset of surveyed sellers, from whom we collected baseline and endline data. We note that markets closer to town especially see decreases in the number of sellers due to treatment, though this does not seem to be explained by different rates of farmers switching to non-local sellers. We speculate that this pattern may be due to their being better alternative business opportunities in less remote areas. We cannot directly observe businesses' opportunity cost of selling seeds, although we do observe that sellers in less remote markets tend to offer a greater variety of agricultural inputs,

[9] Due to logistical difficulties during the main planting season in 2020, we did not collect the total number of sellers present at markets if there were more than 2 seed sellers, and we did not visit all markets. Due to the limit time in March 2020 available before field operations were forced to stop, the field team visited a subset of about 75% of markets in the sample, with this subset driven by logistical considerations and time constraints rather than treatment status or other market characteristics. For the 2020 main planting season, we know for this subset of markets how many sellers were present, or if there are more than 2 sellers, then how many are present among up to two sellers that were randomly selected during baseline activities.

being both more likely to sell other types of vegetable seed as well as being more likely to sell other agricultural inputs.

Table 6: Seller Entry/Exit

	Number of sellers at market
treated	-0.273**
	(0.133)
Observations	495
County FE	Yes
Control Mean	1.58
Treatment Effect (%)	-17.26

This table shows treatment effects on seller entry and exit in the local market. The sample includes markets in the short season in 2020 (approximately seven months after treatment) and the main season in 2021 (approximately thirteen months after treatment). The dependent variable equals the number of sellers present at that market, as observed by secret shoppers. All standard errors are clustered by sublocation.

Other possible stories do not appear to find support in the data. For example, based on our household surveys of local residents (see Table 4), sellers in less remote markets do not appear to have lost more market share due to treatment when compared to less remote areas. We also don't see evidence consistent with lower profit margin per unit in less remote areas (no differences in effects on prices, and no differences in measured profit-per-unit among stayers.)

5.5 Effects on seed quality and prices for uninformed buyers

The data show no detectable effects on visible quality markers or on purity and germination rates of sampled seeds obtained by covert shoppers working for the research team. As Table 7 shows, we see about a 1% increase in packets with all quality markers present. However, these estimates are not statistically significantly different from zero and are far too small to credibly be a driving reason for the increases in yields of the magnitudes that we see. We do not detect effects on purity or germination rate from lab tests.

We interpret this finding to mean that uninformed buyers in treated markets did not obtain higher quality. One reason effects on the uninformed would be even smaller would be if sellers can discriminate between informed and uninformed buyers. In the extreme case in which sellers can perfectly discriminate between informed and uninformed buyers, the uninformed never benefit at all. We also do not see evidence that treatment affected prices for uninformed buyers as revealed by the secret shopper data (Table D9).

6 Discussion

Widespread and effective information campaigns can be difficult and costly to implement. While we pursue an in-person community approach to disseminating information, this design was informed at

Table 7: Quality Markers (all seasons)

	(1) Has all quality markers	(2) purity	(3) germ_rate
treated	0.012	-0.277	-0.539
	(0.028)	(0.274)	(0.753)
Observations	1212	878	878
County FE	Yes	Yes	Yes
Control Mean	0.57	99.85	91.84
Treatment Effect (%)	2.16	-0.28	-0.59

This table shows treatment effects on seed quality offered to secret shoppers posing as uninformed buyers. The sample includes markets in the main season in 2020, the short season in 2020, and the main season in 2021. All standard errors are clustered by sublocation. The dependent variable in column 1 equals 1 if the seed packet has all quality markers present. The dependent variable in column 2 is the estimated percentage of material that are whole maize seeds (in percentage points). The dependent variable in column 3 is percentage of seeds that germinated in the lab (in percentage points). the number of sellers present at that market, as reported by secret shoppers.

least partly by research needs – limiting spillovers, simplicity, highly targeted exposure to information, bundling treatment with data collection activities to save on costs. Using plausible assumptions to estimate the cost of only providing the information campaign (with no data collection costs), we estimate that 25,188 USD was spent to train 12255 residents via face-to-face conversations. This is equivalent to 2.06 USD per person. Valuing harvested maize at 30 USD per 90 kilograms, and taking the point estimate from column one of Table 5, the average benefit is about 16 USD.

Of course, we take such cost-effectiveness figures with a grain of salt. Point estimates come with confidence intervals, and extrapolating will depend on being able to effectively target, especially given the heterogeneous effects in our study sample. One may think of the intervention as lowering the cost of acquiring information. We would expect that other related information interventions may have similar effects and may scale well. For example, reforming the messaging in existing e-verification system to improve accessibility to the information, updating standards for product labeling including standardizing formatting for expiration dates, and a mass-media campaign via newspaper or radio may cost-effectively disseminate information to consumers.

We may also view the cost-effectiveness of treatment through the lens of the toy model of section 3. A sufficiently large increase in the number of informed consumers could have large positive spillovers to uninformed consumers if they induce the local seller to upgrade quality. However, the seller may be induced to quit the market rather than upgrade quality. In this scenario, the spillovers to the uninformed are at best positive only if the uninformed have incorrect beliefs about the quality of seeds that are available, and the seller's exit actually induces uninformed buyers to make a better choice. If uninformed buyers have accurate beliefs about average seed quality in markets, then the seller's exit simply removes an option for uninformed buyers to choose from and is welfare-decreasing. Policies that lower the upgrade cost for sellers (perhaps through training or intervention at the level of the wholesaler) or that reduce the opportunity cost of selling seeds (perhaps through improved access to credit), or that punish the sale of old seed (making the procurement of old seed more costly) could be important for boosting the benefits of improved consumer information.

7 Conclusion

In this paper, we study empirically the effects of reducing information frictions in the hybrid maize seed market in rural Kenya. We evaluate a randomized market-level information campaign to quantify effects for both informed and uninformed buyers in treated market areas. We monitored prices and seed quality that sellers offer for over one year to allow for sellers to respond to the increase in the number of informed buyers.

We show first that farmers stand to gain from receiving information about observable quality markers. The information campaign affected farmers' purchasing decisions and led to gains in maize yields. Second, while improved information caused sellers to exit the market, we do not observe effects on prices or quality among the stayers. This is consistent with a model in which the cost of upgrading from low to high quality are high relatively speaking for sellers, and previously low-quality firms prefer to quit the market rather than be induced to offer high quality.

Our findings suggest that that policies that help improve consumer information may be beneficial to supporting the productivity of small-holder farmers. Reforming existing requirements for information displayed on seed packets for consumers to promote greater understanding about seed certification may be helpful. Yet, the absence of benefits in this study for uninformed consumers points to remaining challenges to overcome the difficult learning and reputation-building environment and their impacts on the market equilibrium.

References

Aker, Jenny and Joshua Blumenstock, "The Economic Impacts of New Technologies in Africa," in Celestin Monga and Justin Lin, eds., *Handbook of Africa and Economics*, Oxford: Oxford University Press, 2015.

Akerlof, George A., "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

Annan, Francis, "Misconduct and Reputation under Imperfect Information," Working Paper, Georgia State University July 2021.

Ashour, Maha, Daniel Orth Gilligan, Jessica Blumer Hoel, and Naureen Iqbal Karachiwalla, "Do Beliefs About Herbicide Quality Correspond with Actual Quality in Local Markets? Evidence from Uganda," *Journal of Development Studies*, June 2019, *55* (6), 1285–1306.

Bai, Jie, "Melons as Lemons: Asymmetric Information, Consumer Learning and Seller Reputation," Technical Report 2021.

Bergquist, Lauren Falcao and Michael Dinerstein, "Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya," *American Economic Review*, December 2020, *110* (12), 3705–3747.

Billings, Lucy; Gilligan Daniel O.; Karachiwalla Naureen Ashour Maha;, "An evaluation of the impact of e-verification on counterfeit agricultural inputs and technology adoption in Uganda," 2015.

Bjorkman-Nyqvist, Martina, Jakob Savensson, and David Yanagizawa-Drott, "Can Good Products Drive Out Bad? Evidence from Local Markets for (Fake?) Antimalarial Medicine in Uganda," CID Working Papers 242, Center for International Development at Harvard University August 2012.

Bold, Tessa, Kayuki C. Kaizzi, Jakob Svensson, and David Yanagizawa-Drott, "Lemon Technologies and Adoption: Measurement, Theory and Evidence from Agricultural Markets in Uganda," *The Quarterly Journal of Economics*, 2017, *132* (3), 1055–1100.

Bronnenberg, Bart J., Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M. Shapiro, "Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium," *The Quarterly Journal of Economics*, 2015, *130* (4), 1669–1726.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins, "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, January 2018, *21* (1), C1–C68.

Dal Bó, Ernesto and Frederico Finan, "At the Intersection: A Review of Institutions in Economic Development," *The Handbook of Economic Development and Institutions*, 2020.

de Brauw, Alan and Berber Kramer, "Improving trust and reciprocity in agricultural input markets: A lab-in-the-field experiment in Bangladesh," Technical Report 2022.

Dranove, David and Ginger Zhe Jin, "Quality Disclosure and Certification: Theory and Practice," *Journal of Economic Literature*, December 2010, *48* (4), 935–963.

Evenson, R E and D Gollin, "Assessing the impact of the green revolution, 1960 to 2000," *Science*, May 2003, *300* (5620), 758–762.

Fabregas, R., M. Kremer, and F. Schilbach, "Realizing the potential of digital development: The case of agricultural advice," *Science*, 12 2019, *366* (6471).

FAO, *2019 Africa Regional Overview of Food Security and Nutrition*, FAO, 2020.

FAO, *2020 Africa regional overview of food security and nutrition*, FAO, 2021.

Gharib, Mariam H., Leah H. Palm-Forster, Travis J. Lybbert, and Kent D. Messer, "Fear of fraud and willingness to pay for hybrid maize seed in Kenya," *Food Policy*, 2021, *102* (C).

Ghassemi-Golezani, Kazem and Sohrab Mamnabi, "SOME PHYSIOLOGICAL RESPONSES AND YIELD OF MAIZE AFFECTED BY SEED AGING AND PRIMING DURATION," *Plant Breeding and Seed Science*, 12 2019, *79*, 63–70.

Gonzalez-Lira, Andres and Ahmed Mushfiq Mobarak, "Slippery Fish: Enforcing Regulation when Agents Learn and Adapt," Working Paper 28610, National Bureau of Economic Research March 2021.

Grubb, Michael D. and Matthew Osborne, "Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock," *American Economic Review*, January 2015, *105* (1), 234–271.

Handel, Benjamin and Joshua Schwartzstein, "Frictions or Mental Gaps: What's Behind the Information We (Don't) Use and When Do We Care?," *Journal of Economic Perspectives*, Winter 2018, *32* (1), 155–178.

Hoel, Jessica, Hope Michelson, Ben Norton, and Victor Manyong, "Misattribution and uncertainty about beliefs prevent learning," 2022.

Holden, Stein T., "Economics of Farm Input Subsidies in Africa," *Annual Review of Resource Economics*, October 2019, *11* (1), 501–522.

Jack, B Kelsey, "Market Inefficiencies and the Adoption of Agricultural Technologies in Developing Countries," Technical Report 2013.

Jin, Ginger Zhe and Phillip Leslie, "Reputational Incentives for Restaurant Hygiene," *American Economic Journal: Microeconomics*, February 2009, *1* (1), 237–267.

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry, " Agricultural Decisions after Relaxing Credit and Risk Constraints *," *The Quarterly Journal of Economics*, 02 2014, *129* (2), 597–652.

Kenya Association of Manufacturers, "The Study to determine the severity of the counterfeit problem in Kenya," Technical Report 2012.

Kenya National Bureau of Statistics, "Kenya Population and Housing Census Results," Technical Report 2019.

Michelson, Hope, Anna Fairbairn, Brenna Ellison, Annemie Maertens, and Victor Manyong, "Misperceived quality: Fertilizer in Tanzania," *Journal of Development Economics*, 2021, *148* (C).

_ , Sydney Gourlay, Travis Lybbert, and Philip Wollburg, "Review: Purchased agricultural input quality and small farms," *Food Policy*, 2023, *116*, 102424.

Muchiri, Joseph, "Fight against fake seeds bearing fruit, says Kephis official," *The Standard*, Jul 2019.

Naritomi, Joana, "Consumers as Tax Auditors," *American Economic Review*, September 2019, *109* (9), 3031–3072.

OECD/FAO, "OECD-FAO Agricultural Outlook," July 2021.

Okinda, Brian, "Alarm as fake seed merchants come up with new ways of selling products," *The Daily Nation*, June 2019.

Shapiro, Carl, "Premiums for High Quality Products as Returns to Reputations," *The Quarterly Journal of Economics*, 1983, *98* (4), 659–679.

Shiferaw, Bekele, Boddupalli M. Prasanna, Jonathan Hellin, and Marianne Bänziger, "Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security," August 2011, *3* (3), 307–327.

Wolinsky, Asher, "Prices as Signals of Product Quality," *Review of Economic Studies*, 1983, *50* (4), 647–658.

A Appendix A: Information campaign details

Figure A1: Treatment #1 Flyer (English Version)

How to check maize seed quality

Quality markers to look for:

1) Clearly displayed company logo together with the seed variety
2) Clearly displayed weight of the seed packet. For example: 2kg, 10kg.
3) Packaging with no holes. The seed packet should not have broken or taped seals that could allow seeds to be removed or tampered with.
4) A printed lot number—this code allows seed packets to be easily traceable
5) A recent packaging or testing date. These dates will tell you if the seeds are old. Older seeds may germinate poorly.
6) Seeds should not be split or broken, and the coloring on the seeds should not come off easily
7) A KEPHIS sticker:

- Check that the SMS code is valid
- Check that the variety, lot number, and testing date match the packaging

Valid	Not valid
OK Monsanto DK8031 Species: Zea mays Variety:DK8031 Lot No: 18-23463HP Class: C1G Testing Date: Jan/2019 More: ghub.ai/zH9d	No 219823200694 IS NOT A VALID CODE. Check and send correct code. The seed may not be genuine. Call 0709891000 or ke@mpedigree.net. More: ghub.ai/X2XD
	No: 261710114026 was a Valid Code BUT was used on 2019-03-14 14:23:56 by 7151****8. MPedigree service. hub.goldkeys.net More: ghub.ai/zzfc

Figure A2: Treatment #2 Flyer (English Version)

If you are concerned about the quality of your seeds, you may contact the following:

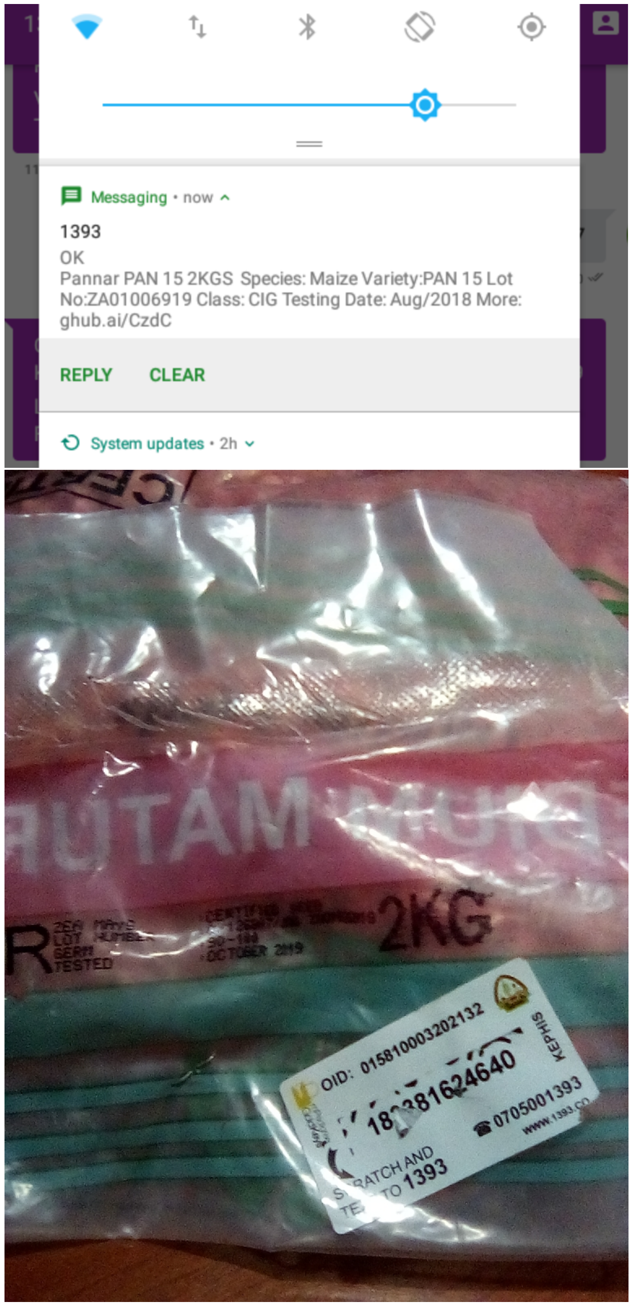
- Kenya Plant Health Inspectorate Service (KEPHIS)
  Phone Number: 020-3597209
- Anti-Counterfeit Authority (ACA)
  Phone Number: 020 2280111
- Your assistant chief

Other Information:

- Not all seeds are of the same quality. It is important to discuss with friends, family, and neighbors the quality of seeds that you purchase. This can help them avoid seeds that may be lower quality.
- When buying and planting maize seeds, it can be helpful to ask the seller for a receipt and keep the empty packet after planting. In case there is any issue with the quality of the seeds, you can refer to these items to confirm the type of seed, as well as the time and place where you purchased the seeds.

Figure A3: Maize variety with (left) and without (right) quality marks

Figure A4: E-verification



B Appendix B: Model

B.1 Model details

In section 3, the firm's problem is as follows, where we write x as a function of the firm's choice of price p and quality mix q:

$$Max_{p,q}(p - c - d\frac{q}{1 - \theta + q\theta}) * x(p, q) - F$$

Accounting for how consumers respond to the firm's choice of p and q, the firm's problem becomes:

$$Max_q(p - c - d\frac{q}{1 - \theta + q\theta})(1 - \theta + \theta q)N - F$$

This simplifies to:

$$Max_q(p - c)(1 - \theta + \theta q)N - dqN - F$$

Taking derivatives with respect to q, we get:

$$\frac{\partial}{\partial q}(\cdot) = (p - c)\theta N - dN$$
$$= (p - c)\theta - d$$

From this, we can define conditions under which the firm chooses $q = 1$ or $q = 0$:

$$q^* = 1 \text{ if } (p - c)\theta - d \geq 0$$
$$q^* = 0 \text{ if } (p - c)\theta - d \leq 0$$

Any level of q^* would be optimal in the edge case where we have exact equality.

The firm's maximized profit is:

$$(p - c)(1 - \theta + \theta q^*)N - dq^*N - F$$

Taking a derivative with respect to θ, we see that profit is always at least weakly decreasing in the choice of quality, and strictly decreasing if the optimal quality choice is not 1:

$$\frac{\partial \pi^*}{\partial \theta} = -(1 - q^*)(p - c)N \leq 0$$

Taking stock, the firm will be incentivized to choose high quality if $\theta > \frac{d}{p-c}$. And the firm will choose to exit the market if $\theta > 1 - \frac{F+\pi_0}{(p-c)N}$. As θ rises, whether a firm offering low quality will upgrade quality or quit the market depends on which threshold is lower. If we combine these inequalities, we can see that the firm will opt to upgrade quality over quitting if:

$$d < (p - c) - \frac{F + \pi_0}{N}$$

That is, the firm will upgrade quality if there are enough informed consumers, only if the upgrade cost is not too high relative to a function of the profit margin and average fixed cost (in fact, the average profit per unit sold of low quality seed).

This inequality is critical for the spillover effects of improved consumer information. In this toy model, there are three different types of consumers that experience different changes in utility in the case where the inequality holds, and where it does not hold.

Case 1: As θ increases, it induces firms to switch from $q = 0$ to $q = 1$.

- The always-informed: go from getting $q - P_0 - S$ to $1 - p$, as they gain the option to buy high quality seeds locally.
- The newly-informed: go from getting -p utility to 1-p utility, as they switch to buying local high quality seeds
- The never-informed: go from getting -p utility to 1-p utility

Here, improved consumer information benefits everyone.

Case 2: As θ increases, it induces firms to exit.

- The always-informed: Have no change in utility, as they left the local market to purchase seeds even at baseline
- The newly-informed: go from getting -p utility to $1 - P_0 - S$ utility, as they switch to buying high quality seeds from the outside option
- The never-informed: go from getting -p utility to $\bar{q} - P - 0 - S$ utility.

Here, the always informed do not benefit from improved consumer information, and the never-informed perceive that they are worse off. However, it is possible that they do benefit if their beliefs about quality are incorrect, and the removal of the local low-quality option actually serves to remove a tempting but worse option.

B.2 Conditions for equilibrium

In section 3 we consider the equilibrium in which sellers set price low enough to entice uninformed buyers and informed buyers who are offered high quality to buy locally. Below, we discuss the alternative cases where price is set higher or lower.

Alternative case #1: the firm could choose a lower price to keep even informed consumers who get low quality. Those consumers get $1 - P_0 - S$ in the outside option and $-p$ at the local firm. Therefore, this informed consumer who gets low quality will stay if: $-p \geq 1 - P_0 - S$, or $p < P_0 + S - 1$.

The optimal quality in this case will be q=0. The firm's profits are:

$$(p - c)N - \frac{1}{2}b\theta - F$$

Price cannot drop below c or else the firm is sure to exit the market. Thus we can rule out this scenario if buyer's valuation of a high quality product is large enough. We also note that we see a good number of packets with all quality markers, suggesting that we generally do not observe an equilibrium with $q = 0$.

Alternative case #2: the firm chooses a higher price $P_0 + S$ to keep only the informed buyers who get high quality (uninformed buyers feel they are better off getting the higher quality in the outside option). Note that as $\hat{q} - \bar{q}$ goes to zero, this becomes a sub-optimal choice for any values for the parameters, since the firm loses a large segment of customers (the uninformed customers) for a vanishingly small gain in price per unit, and must stock exclusively high quality for all of the informed customers to keep them buying locally and prevent the outside option from being the more attractive option. We note also that the share of uninformed is generally large in our sample. Among control group households, and using either knowledge of expiration date or knowledge of the KEPHIS sticker to mean that the consumers is "informed", we estimate that 84% are uninformed.

C Appendix C: Spillovers Effects And Effects of Baseline Surveys

C.1 Spillover Effects

In this subsection, we examine whether nearby treatment sites had effects on the major outcomes. To do so, in the main estimating equation to recover treatment effects, we control for the number of study sites within 2, 4, or 6 kilometers, and we also include the number of treated sites within that distance.

In Table C1 we examine possible spillovers on household knowledge.

Table C1: Households knowledge (spillovers)

Panel A: Spillovers within 2km

	(1) no. quality marks known	(2) e-verif known	(3) expiration known
Number of treatment sites within 2km	-0.035	0.022	-0.006
	(0.059)	(0.020)	(0.015)
Number of control sites within 2km	0.046	0.035	0.000
	(0.061)	(0.022)	(0.015)
Observations	2425	2425	2425
County FE	Yes	Yes	Yes
HH Baseline Controls	Yes	Yes	Yes

Panel B: Spillovers within 4km

	(1) no. quality marks known	(2) e-verif known	(3) expiration known
Number of treatment sites within 4km	-0.020	0.009	-0.012
	(0.033)	(0.011)	(0.008)
Number of control sites within 4km	0.035	0.012	0.007
	(0.028)	(0.009)	(0.007)
Observations	2425	2425	2425
County FE	Yes	Yes	Yes
HH Baseline Controls	Yes	Yes	Yes

Panel C: Spillovers within 6km

	(1) no. quality marks known	(2) e-verif known	(3) expiration known
Number of treatment sites within 6km	0.016	0.012*	0.001
	(0.022)	(0.007)	(0.007)
Number of control sites within 6km	0.036*	0.011**	0.007
	(0.019)	(0.005)	(0.005)
Observations	2425	2425	2425
County FE	Yes	Yes	Yes
HH Baseline Controls	Yes	Yes	Yes

We examine in Table C2 whether treatment may have had other spillover effects. First, we examine effects on firm entry and exit. We examine in Table Table C3 possible spillover effects on quality markers and lab tested quality.

C.2 Effects of Baseline Surveys

We compare control sites (which received baseline surveys but no information campaign) with pure control sites (which received no baseline activities of any kind and were visited for the first time in March to April 2021). For data from the market audit survey, we have data for the LGS2021 planting season for both control markets and pure control markets. Table C4 shows no statistically significant differences in the numbers of sellers or in the prevalence of quality markers in control markets as compared to pure control markets.

Table C2: Seller Entry/Exit Spillovers (spillovers)

Panel A: Spillovers within 2km

	(1) Number of sellers at market
Number of treatment sites within 2km	-0.184
	(0.156)
Number of control sites within 2km	0.053
	(0.206)
Observations	371
County FE	Yes
HH Baseline Controls	

Panel B: Spillovers within 4km

	(1) Number of sellers at market
Number of treatment sites within 4km	-0.101
	(0.126)
Number of control sites within 4km	-0.072
	(0.077)
Observations	371
County FE	Yes
HH Baseline Controls	

Panel C: Spillovers within 6km

	(1) Number of sellers at market
Number of treatment sites within 6km	-0.057
	(0.067)
Number of control sites within 6km	-0.038
	(0.061)
Observations	371
County FE	Yes
HH Baseline Controls	

Table C3: Quality Markers and Lab Tests Spillovers (spillovers)

Panel A: Spillovers within 2km

	(1)	(2)	(3)
	Has all quality markers	purity	germ_rate
Number of treatment sites within 2km	-0.036	0.014	-1.372
	(0.050)	(0.019)	(1.301)
Number of control sites within 2km	-0.016	-0.008	0.019
	(0.044)	(0.017)	(1.171)
Observations	662	484	484
County FE	Yes	Yes	Yes
HH Baseline Controls			

Panel B: Spillovers within 4km

	(1)	(2)	(3)
	Has all quality markers	purity	germ_rate
Number of treatment sites within 4km	-0.017	-0.009	-0.060
	(0.024)	(0.009)	(0.645)
Number of control sites within 4km	0.005	0.003	0.080
	(0.021)	(0.008)	(0.557)
Observations	662	484	484
County FE	Yes	Yes	Yes
HH Baseline Controls			

Panel C: Spillovers within 6km

	(1)	(2)	(3)
	Has all quality markers	purity	germ_rate
Number of treatment sites within 6km	-0.021	-0.000	0.584
	(0.015)	(0.006)	(0.399)
Number of control sites within 6km	0.006	-0.001	0.472
	(0.014)	(0.005)	(0.360)
Observations	662	484	484
County FE	Yes	Yes	Yes
HH Baseline Controls			

Table C4: Control vs Pure Control: LGS2021 Firm Entry/Exit

	(1)
	Number of sellers at market
Baseline Surveys Administered	0.098
	(0.157)
Observations	195
County FE	Yes
Control Mean	1.771

Table C5: Control vs Pure Control: LGS2021 Quality Markers

	(1) Has all quality markers	(2) Has valid KEPHIS sticker	(3) Lot number present	(4) Has expiration date	(5) Has valid expira
Baseline Surveys Administered	-0.014	-0.025	-0.008	-0.020	-0.030
	(0.058)	(0.021)	(0.008)	(0.057)	(0.057)
Observations	258	254	265	266	266
County FE	Yes	Yes	Yes	Yes	Yes
Control Mean	0.616	0.966	0.994	0.662	0.637

	(1) purity	(2) germ_rate
Baseline Surveys Administered	0.054	3.154*
	(0.039)	(1.818)
Observations	165	165
County FE	Yes	Yes
Control Mean	99.86	91.90

Table C6: Control vs Pure Control: LGS2021 Prices

	(1)
	Price paid for 2kg hybrid
gotbaseline	-2.635
	(7.227)
Observations	266
County FE	Yes
Control Mean	483.2

D Appendix D: Extra Tables and Figures

Table D1: Treatment Delivery

	Flyers distributed	Residents trained directly
Avg number per site	173.5	81.7
% of local customers	19.7%	9.1%

Table D2: Baseline Balance

Variable	Control Mean	Treatment Mean	Difference
Market Area Population	143.5	138.8	-4.660
			(9.695)
No. Seed Sellers	4.103	4.519	0.416
			(0.643)
HH Head Gender (1 = Male)	0.704	0.710	0.00600
			(0.020)
HH Head Age	49.97	50.64	0.672
			(0.642)
Completed Primary School	0.623	0.583	-0.040*
			(0.023)
Completed Secondary School	0.265	0.225	-0.040**
			(0.019)
Home Quality Index	1.438	1.439	0.00200
			(0.049)
Acres Planted (2019 main season)	1.157	1.130	-0.0280
			(0.051)
Germination Rate (2019 main season)	82.04	82.28	0.236
			(0.851)
Hybrid Maize Yield (Kgs/acre 2019 main season)	910.8	876.0	-34.81
			(38.570)
Joint F test – p-value = .562			

The sample includes 2373 households surveyed at baseline. P-value for joint test for differences between treatment groups is 0.44. For all tests for differences between the treatment group and the control group, we control for dummies for each county – which is the level at which site selection was stratified – and we calculate standard errors clustered at the sub-location level, which is the unit of randomization.

Table D3: Quality markers and seed purity

	Purity	Purity	Purity	Purity	Purity	Purity
Has all quality markers	0.0					
	(0.0)					
Has valid KEPHIS sticker		-0.0				
		(0.0)				
Lot number present			-0.0			
			(0.1)			
Has expiration date				0.0**		
				(0.0)		
Has valid expiration date					0.0*	
					(0.0)	
No holes						-0.0
						(0.1)
Constant	99.8***	99.9***	99.9***	99.8***	99.8***	99.9***
	(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(0.1)
Observations	467	464	467	468	468	468

The dependent variable is seed purity, measured in percentage points. 100 percent purity indicates that a sample from the seed packet contained only whole seeds.

Table D4: Quality markers and germination rate

	Germination Rate	Germination Rate	Germination Rate	Germination Rate	Germination Rate	Germination Rate
Has all quality markers	5.2***					
	(1.1)					
Has valid KEPHIS sticker		0.9				
		(2.1)				
Lot number present			-2.6			
			(4.4)			
Has expiration date				3.2**		
				(1.4)		
Has valid expiration date					6.3***	
					(1.2)	
No holes						-4.3
						(4.4)
Constant	88.1***	90.9***	94.3***	89.0***	87.0***	96.0***
	(1.0)	(2.0)	(4.3)	(1.3)	(1.0)	(4.3)
Observations	467	464	467	468	468	468

The dependent variable is germination rate, measured in percentage points. 100 germination rate indicates that, in following standard procedures at KEPHIS facilities, all sampled seeds from the seed packet germinated under ideal lab-controlled conditions.

Table D5: Households knowledge (by education)

	(1) no. quality marks known	(2) e-verif known	(3) expiration known
Treat x No Primary Educ	-0.017	0.051***	-0.003
	(0.049)	(0.015)	(0.013)
Treat x At Least Primary Educ	0.418***	0.171***	0.053***
	(0.052)	(0.017)	(0.013)
Observations	4456	4456	4456
Control Mean	0.56	0.13	0.07

This table shows treatment effects on household knowledge about quality markers. The dependent variables are (1) the number of quality markers named by the respondent, (2) whether e-verification was named as a way to verify seed quality (=1 if yes), and (3) whether expiration date was named (=1 if yes). In the estimating equation, a binary treatment indicator is interacted with either a dummy indicating that the household head has completed primary school, or a dummy indicating that the household has not completed primary school. All standard errors are clustered by sublocation.

Table D6: Household complementary inputs (LGS2020/SGS2020)

	(1) Used Fertilizer
treated	-0.00
	(0.01)
Observations	2871
County FE	Yes
HH Baseline Controls	Yes
Control Mean	0.93
Treatment Effect (%)	-0.24

This table shows treatment effects on household use of fertilizer. Results include households in both the main season in 2020 and the short season in 2020. The dependent variable equals 1 if the household indicated that they used fertilizer (at all) for maize during that planting season. Standard errors are clustered by sublocation.

Table D7: Prices for hybrid maize seed (household self-reported)

	(1) Avg price per kg	(2) Avg price per kg from local market	(3) Avg price per kg from elsewhere
treated	2.723	-0.330	3.736
	(1.831)	(2.081)	(2.538)
Observations	2217	1162	1028
County FE	Yes	Yes	Yes
HH Baseline Controls	Yes	Yes	Yes
Control Mean	219.14	223.85	219.13
Treatment Effect (%)	1.24	-0.15	1.71

Table D8: Price dispersion for hybrid maize seed (household self-reported)

	(1) SD of price per kg in sublocation	(2) Range of price per kg in sublocation
treated	-1.679	-3.274
	(1.898)	(3.779)
Observations	218	251
County FE	Yes	Yes
HH Baseline Controls	Yes	Yes
Control Mean	15.95	31.44
Treatment Effect (%)	-10.53	-10.41

Table D9: Prices for hybrid maize seed (from market audit; All Seasons)

	(1) Price paid for 2kg hybrid
treated	0.091
	(5.102)
Observations	693
County FE	Yes
HH Baseline Controls	Yes
Control Mean	480.30
Treatment Effect (%)	0.02

Table D10: Treatment effects on choice to buy seeds at local market: robustness checks

	(1) Bought hybrid at local market	(2) Bought hybrid at local market	(3) Bought hybrid at local market	(4) Bought hybrid at local market	(5) Bought hybrid at local market	(6) Bought hybrid at local market	(7) Bought hybrid at local market
treated	-0.053**	-0.048*	-0.051*	-0.053**	-0.053**	-0.054**	-0.045*
	(0.025)	(0.027)	(0.026)	(0.025)	(0.025)	(0.025)	(0.025)
Observations	4086	4074	3956	4086	4086	4086	3928
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HH Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	0.32	0.32	0.32	0.32	0.32	0.32	0.32
Treatment Effect (%)	-16.71	-15.14	-15.94	-16.71	-16.63	-17.10	-14.09
Specification	Baseline specification	Household weights	Only hybrid seed users	Winsorized at 95th percentile	No adjustment for baseline imbalance	Controlling for baseline seed brands used	Double Lasso (weighted)

Table D11: Treatment effects on kgs harvested: robustness checks

	(1) Kgs Harvested	(2) Kgs Harvested	(3) Kgs Harvested	(4) Kgs Harvested	(5) Log(kgs harvested)	(6) Kgs Harvested	(7) Kgs Harvested	(8) Kgs Harvested
treated	57.75**	57.98**	59.93**	41.21**	0.16***	54.21*	58.11**	59.57**
	(29.05)	(29.11)	(29.50)	(19.69)	(0.06)	(29.66)	(29.09)	(28.03)
Observations	3807	3795	3698	3807	2447	3807	3807	3670
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HH Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	551.59	552.12	561.06	462.84	869.26	551.59	551.59	562.25
Treatment Effect (%)	10.47	10.50	10.68	8.90		9.83	10.53	10.60
Specification	Baseline specification	Household weights	Only hybrid seed users	Winsorized at 95th percentile	Log transform	No adjustment for baseline imbalance	Controlling for baseline seed brands used	Double Lasso (weighted)

Table D12: Treatment effects on maize yield: robustness checks

	(1) Yield	(2) Yield	(3) Yield	(4) Yield	(5) Log(yield)	(6) Yield	(7) Yield	(8) Yield
treated	54.84**	44.28	59.40**	59.72**	0.13***	50.42*	54.51**	59.15**
	(27.39)	(31.55)	(27.95)	(23.75)	(0.05)	(27.94)	(27.31)	(28.48)
Observations	2443	2433	2370	2443	2443	2443	2443	2355
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HH Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	846.88	847.04	853.66	806.44	846.88	846.88	846.88	853.33
Treatment Effect (%)	6.48	5.23	6.96	7.41		5.95	6.44	6.93
Specification	Baseline specification	Household weights	Only hybrid seed users	Winsorized at 95th percentile	Log transform	No adjustment for baseline imbalance	Controlling for baseline seed brands used	Double Lasso (weighted)

Table D13: Treatment effects, excluding households that purchased seeds prior to treatment

	(1) Kgs harvested	(2) Yield
treated	65.92**	52.91*
	(30.13)	(29.38)
Observations	3470	2112
County FE	Yes	Yes
HH Baseline Controls	Yes	Yes
Control Mean	508.10	840.21
Treatment Effect (%)	12.97	6.30

Table D14: Treatment effects, by level of competition at baseline

	(1) Kgs harvested	(2) Yield
Treated x 1 Baseline Seller	73.76	38.70
	(47.76)	(42.36)
Treated x 2+ Baseline Sellers	46.52	63.42**
	(31.86)	(30.14)
Observations	3807	2443
County FE	Yes	Yes
HH Baseline Controls	Yes	Yes
Control Mean	551.59	846.88

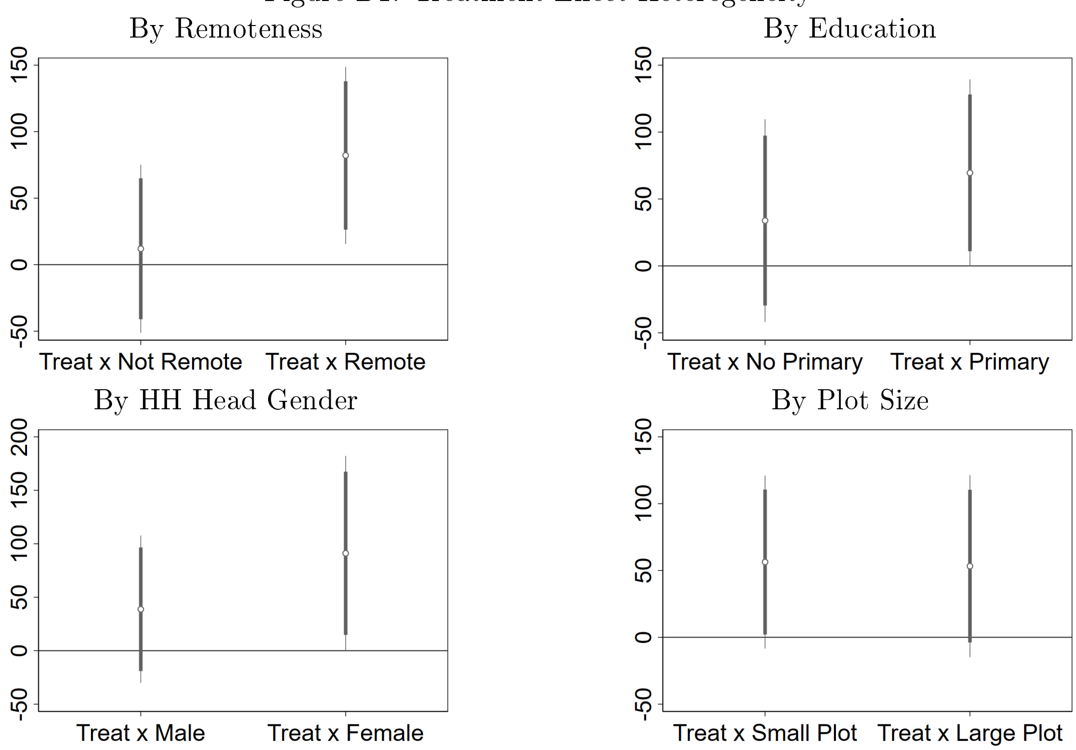
Table D15: Treatment effects, excluding households with close personal relation to seller

	(1) Kgs harvested	(2) Yield
treated	124.10**	61.10**
	(48.65)	(30.35)
Observations	2092	2086
County FE	Yes	Yes
HH Baseline Controls	Yes	Yes
Control Mean	854.24	847.00
Treatment Effect (%)	14.53	7.21

Table D16: Treatment effects (double machine learning)

	(1) Kgs harvested	(2) Yield
treated	47.44**	44.84**
	(20.73)	(22.14)
Observations	2350	2344
County FE	Yes	Yes
HH Baseline Controls	Double ML	Double ML
Control Mean	700.26	808.10
Treatment Effect (%)	6.78	5.55

Figure D1: Treatment Effect Heterogeneity



This figure illustrates heterogeneity in treatment effect effects on maize yield. The top-left figure shows estimates by remoteness, where remote is defined as above the median distance to the county capital. The top right figure shows estimates by gender of the household head. The bottom left figure shows estimates by education status of the household head, as measured by whether or not the household head had completed primary school. The bottom right figure shows estimates by plot size, where household observations are split by the median number of acres owned.