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

Eric Hsu[a,1] Anne Wambugu[b]

This version: May 2024

Abstract

Hybrid seeds are an important technology for climate change adaptation. However, their quality is difficult for buyers to observe, an information friction that can impact the functioning of rural markets as an effective way to distribute seeds. In this paper, we first document that substandard seeds are common in rural Kenyan markets. We then study how increasing the number of informed buyers–who can discern product quality–affects market outcomes over time. To do so, we implement a market-level intervention, randomizing rural markets into a community-wide information campaign that trains farmers to identify quality-verified hybrid maize seeds. The intervention improved knowledge, affected seed purchases, and increased maize production. Impacts were heterogeneous—more educated farmers and areas with more substandard seeds benefited more. Meanwhile, sellers exited in response to treatment, but did not adjust prices or quality. These patterns can be explained by a simple model in which informed buyers can detect quality and switch sellers if needed to obtain higher quality. The presence of more informed buyers hurts profits of local sellers and can induce them to exit rather than offer higher quality if upgrading quality is too costly. The impacts on access to high quality seeds are far lower in this world relative to a world where firms are induced to upgrade quality. Taken together, the findings document new stylized facts and provide evidence relevant for boosting yields of a staple crop. More generally, they provide lessons on the role of improved consumer information in disciplining firms in low information environments.

JEL Codes: D18, D22, D83, K42, L15, O13, Q1

Word count: 12545

[a] Yale University [b] Strathmore University. [1] Corresponding Author. Yale Research Initiative on Innovation and Scale, Whitney and Betty MacMillan Center for International and Area Studies, Yale University, United States of America. Email address: eric.hsu@yale.edu. *We thank Francis Annan, Susanna Berkouwer, Stephanie Bonds, Benjamin Faber, Lauren Falcao Bergquist, Jishnu Das, Kyle Emerick, Frederico Finan, Benjamin Handel, Jessica Hoel, Supreet Kaur, Kei Kawai, Amit Khandelwal, 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, IZA/FCDO G2LM-LIC Research Conference, Centre for the Study of African Economies Annual Conference, and Yale Development Lunch 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

The adoption of hybrid seeds is a key strategy for climate change adaptation, particularly for some of the world's most vulnerable populations threatened by food insecurity (World Bank, IFC, and MIGA, 2016; Kala et al., 2023) Like many technologies – including health products, home appliances, and other improved agricultural inputs – the benefits depend crucially on products not being expired, damaged, adulterated, or otherwise being of substandard quality.[1] However, when quality depends on unobserved properties of the technology, verifying product quality can be extremely challenging for users. This inability to observe quality can lead to a market breakdown in which sellers find it unprofitable to offer high quality versions of the products (Akerlof, 1970). Markets in low and middle income countries (LMICs) may be particularly vulnerable when counteracting mechanisms function imperfectly (Dranove and Jin, 2010; Abate et al., 2021). For example, enforcement of government quality standards and other industry regulations may be undermined by lower state capacity in these settings (Dal Bó and Finan, 2020).

In this paper, we study the market-level effects of policies that improve consumers' ability to detect product quality in the hybrid maize (corn) seed market in rural Kenya. Several features of this setting may make it difficult for buyers to learn 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. First, infrequency of purchase limits the speed at which consumers can learn about seed quality. Typical crop cycles limit farmers to only observing one or two harvests per year. Second, even after harvest the performance of 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 further impeding learning.[2] Third, while government regulations exist to establish minimum quality standards for seeds, enforcement is limited in this setting. For example, regulators conduct testing after manufacture but prior to certification, but they have limited capacity to monitor quality along the distribution chain all the way to the end-user. As a result, there have been widespread concerns in recent years about the prevalence of seeds that are not quality verified, including expired, adulterated, and poorly stored seeds (Okinda, 2019).

Improving farmers' ability to access high quality seeds is particularly important in this setting. High quality maize seeds are a productive input critical for maintaining food security and battling 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, where over 20% of total calories are supplied by maize (Shiferaw et al., 2011). In the

[1] For instance, the proliferation of substandard medicines, electronics, and agricultural products have been documented to be a significant issue in Kenya (Kenya Association of Manufacturers, 2012) and around the world especially in lower income countries (Harris and Morris, 2009; Schneider et al., 2020; Michelson et al., 2023)

[2] Cross-sectional data on yields from this paper suggests that quality-verified seeds could increase yields by less than 0.1 standard deviations. Panel data from the Ugandan LSMS surveys suggests that a 10% increase in yields would be equivalent to about a 0.14 increase in standard deviation terms, where the relevant variance in yields is computed as the variance of the residual in a household fixed effects regression controlling for plot size (Bold et al., 2017)

East Africa region, food security is a widespread and chronic issue – over 30% of the population experience severe food insecurity, indicating a lack of adequate access to enough safe and nutritious food for healthy development (FAO, 2020, 2021). As a result, 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 following 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, community-wide trainings were conducted to enable farmers to better detect quality-verified hybrid seeds. A market-level intervention makes it possible to identify effects of improved information on the behavior of local market participants, including seed sellers, consumers who became informed due to treatment, and uninformed consumers in treated communities. We examine whether and how informed buyers can achieve improved agricultural outcomes; we also examine how sellers respond to the information treatment in their market entry, price setting, and quality setting decisions. These adjustments made by sellers determine to what extent uninformed buyers in treated communities may also experience benefits, through improved access to seeds of better quality. A series of surveys and secret shopper activities track outcomes for buyers and sellers–including knowledge, seed sales and purchases, and agricultural outcomes–for over one year after treatment. The extended follow-up period of over one year makes it possible to track possible convergence to a new equilibrium over the course of three planting seasons following treatment. The sample includes 104 pure control sites, which allows us to test whether baseline survey activities could have influenced outcomes.

The analysis produces three main findings. First, the analysis confirms that observable quality markers predict seed quality. Seed packets missing one or more quality markers are common and were found to have lower lab-tested germination rates, one key measure of seed quality.[3] While germination rate is far from the only aspect of seed quality that is relevant for maize yields, this finding nonetheless indicates 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 these observable quality markers is low at baseline.

Second, the information treatment improved buyers' ability to detect seed quality. Treated buyers had greater knowledge of observable quality markers and were more likely to successfully use detection techniques. Treated buyers frequently reported that the information that was provided affected their purchasing decisions—both which packets they bought and which seller they bought from. Consistent with these accounts and with a simple model of consumer search behavior, we see 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, suggesting that treated farmers obtained higher quality seeds and this led to increased productivity. More remote farmers benefited especially (consistent with data showing larger differences in seed quality between verified and unverified packets in such areas) as did more

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

educated farmers (consistent with their demonstrating substantially greater information retention).

Third, the intervention caused sellers to exit the market for hybrid maize seed in the subsequent planting seasons, as revealed by data from secret shoppers. Sellers responded minimally in the first planting season, about one month after treatment. By the second and third planting seasons–about 7 and 13 months after treatment–approximately one in six sellers on net exited the market due to treatment. Among sellers that remained in the markets, we do not detect treatment effects on seed quality offered to uninformed buyers, nor do we detect effects on price or price dispersion. While individuals who were not initially treated could have benefited from information spillovers, these findings show it is unlikely that spillovers occurred through any equilibrium effects on price adjustments or quality upgrading by sellers.

These patterns can be explained by a simple model in which informed farmers can conduct a sequential search for high-quality seeds–possibly leaving the local market to do so. As the number of informed farmers increases, this hurts the profits of local firms that do not offer high quality, as they lose market share. If local firms do not upgrade quality, they lose market share as more consumers become informed. If the cost to the firm of upgrading quality is low enough, then firms can be induced to upgrade quality with a sufficiently large increase in the number of informed consumers. However, if the upgrade cost is too high, then firms will instead prefer to exit the market for seeds in response to improved consumer information. The empirical findings showing strong firm exit effects point to the need for further work to identify reasons why firms may face high costs to upgrading quality.

Taken together, the results document new stylized facts and provide policy-relevant evidence that helps promote the adoption of an important technology. 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 both with accounts of widespread counterfeiting (Kenya Association of Manufacturers, 2012) and with other problems with quality in the supply chain. Meanwhile, uncertainty about quality of inputs could have additional negative impacts by depressing investment in other complementary inputs (Bulte et al., 2023).

Our paper contributes to this literature by evaluating an intervention to improve consumer information about quality 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 offers important descriptive data on the quality of agricultural inputs (Ashour et al., 2019; Bold et al., 2017; Gharib et al., 2021; Kenya Association of Manufacturers, 2012; Michelson et al., 2021)), particularly on the relationship between observables and seed quality.

The paper also adds 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; Leland, 1979; Shapiro, 1983; Wolinsky, 1983). A number of studies have tried to test theoretical predictions about quality provision in real-world settings when buyers have imperfect information about quality (Bai, 2021; Barahona et al., 2023; Bennett and Yin, 2019; Bjorkman-Nyqvist et al., 2012; Hotz and Xiao, 2011; Hui et al., 2022; Jensen and Miller, 2018; Jin and Leslie, 2009). We contribute to this literature by empirically examining seller and buyer responses to improved information in a high-stakes and policy-relevant context, using an experimental evaluation closely tied to a common government policy (mandatory product certification). Differently from many of these studies, we do not find evidence of firms adjusting on the quality margin. This paper also relates to work on consumer misinformation, costly consumer mistakes, and incorrect beliefs (Bronnenberg et al., 2015; Grubb and Osborne, 2015; Handel and Schwartzstein, 2018; Hoel et al., 2022). 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 product certification. Several papers examine how consumer information or monitoring by consumers (Gonzalez-Lira and Mobarak, 2021; Annan, 2021; Naritomi, 2019) can complement or replace direct enforcement efforts (Duflo et al., 2018). Different from previous work, we find that in some settings consumers may be limited in their ability to induce firms to comply with regulations.

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 a major source of calories and is an important crop especially 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, particularly in the face of changing climate conditions. In the case of maize, yields in lower income countries are a small fraction of average yields seen in OECD countries (OECD/FAO, 2021). As a result, increasing access to improved agricultural inputs such as high yielding seed varieties and fertilizers have been one focus among policymakers in recent decades (Evenson and Gollin, 2003). This has become increasingly important in the face of the effects of climate change, and high-yielding and resilient hybrid seed varieties is a leading technology to help farmers adapt.

However, even when a farmer believes they have adopted an improved variety of maize, its quality may be substandard. 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 as they can, for example, be adulterated, expired, or improperly stored (Ashour et al., 2019; Bold et al., 2017; Michuda et al., 2022; 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. All seeds must be certified by KEPHIS before sale, and all seed sellers must register with KEPHIS. While certification imposes some requirements such as field inspections during seed manufacturing, other requirements must be met after manufacture. Importantly, seeds must test with over 90% germination rate to be certified, meaning that under ideal temperature and moisture conditions randomly sampled seeds from each production lot must develop normally at high rates. Without passing required checks to obtain certification, seeds cannot legally be distributed to wholesalers and retailers for sale. Concerns about the prevalence of uncertified seeds led to a new initiative in 2018 to mandate e-verification for certified seeds, similar to verification schemes implemented in other settings (e.g. agricultural inputs in Uganda (Ashour, 2015; Okwakol, 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. Besides offering a binary indicator that the seeds were certified, the new e-verification scheme allowed for an additional method to obtain and verify printed information–namely receiving key information about the seeds such as the packaging date 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 to limit fraudulent reuse. Subsequent uses would result in a message indicating that the code is valid but also alerting the user that it had been used before.[4]

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." (Ngila, 2019) Counterfeits may not be the only possible cause of low-quality seeds, however. Other reasons include poor storage or selling seeds after expiration or too long after testing to ensure high quality, though in practice (including in this paper's analyses) it can be challenging to distinguish empirically among these possibilities. This is consistent with academic research on maize seed performance after storage under less-than-ideal conditions which degrades seed performance (Ghassemi-Golezani and Mamnabi, 2019). 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 can allow for lower-quality seed packets to 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,

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

allowing for counterfeits or other non-quality-verified products to enter the supply chain. While the regulator tests seeds at the certification step taking place after manufacture but before distribution, there is limited capacity to monitor seed quality at the end of the distribution chain at which farmers purchase seeds. In practice, small retailers receive little attention from monitoring efforts, which have focused on larger distributors (e.g. Obura, 2013) though these efforts do not appear to have yielded punitive sanctions in court.

We study retail shops in a study area that includes four counties in western Kenya - 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). Over 90% of farmers in this study's sample believed at baseline that maize seed quality is an issue that sometimes cause poor harvests, on average reporting that they believed between 10 and 20 percent of packets are problematic.

3 A Simple Model With Informed and Uninformed Consumers

To structure our thinking about how improved consumer information may affect the behavior of sellers and consumers, we consider a simple one-period model. In its assumptions, the model matches features observed in rural markets for hybrid maize seed on our study setting. We assume one local seller that sets price and quality levels for a single good. Quality is assumed to be binary – either high or low – representing quality-verified seeds as opposed to seeds that are not quality-verified in the real world. The local market includes many informed consumers but also many uninformed consumers, with each consumer seeking to buy one unit of the good. The necessity to buy the good corresponds to very high take-up rates of hybrid seeds in our study area, with 89% of farmers in our sample using hybrid seeds in the main planting season. Consumers can buy from the local seller or alternatively they can buy from another seller that is outside the local market. We focus our attention on an equilibrium in which informed buyers (who are able to observe quality before making a purchasing decision) opt to leave the local market if quality is observed to be low but buy from the local seller if quality is observed to be high. The model illustrates that when the proportion of informed buyers increase, a seller that initially offers only low-quality products will be induced to either upgrade the quality they offer or to exit the market. This is because as more and more of the local consumers become informed, the seller's profit when offering low quality approaches zero as it loses market share. If it is able to make a larger (and positive) profit by switching to selling high-quality products, then it will be induced to do so when the market has enough informed buyers. Otherwise, it will exit the market and accept zero profit.

We will conclude in this section that when the cost to the firm of upgrading quality is small enough, then improving consumer information could help even consumers that remain uninformed. This happens because when sellers improve the quality that they offer, this causes all local buyers to obtain high-quality seeds when they shop locally, including those who are unable to observe quality before purchase. However, if the upgrade cost is too high, then the presence of even very many informed buyers may not induce sellers to upgrade quality; they will exit the market instead, and as a result uninformed consumers do not gain better access to high-quality seeds. (Dranove and Jin, 2010) provides a summary of the theoretical literature on unraveling in markets with information asymmetry, in which sellers may voluntarily divulge credible information about the quality of products they sell. We abstract away from the possibility that consumer information could be endogenous. We also abstract away from competitive forces in the local market by assuming only a single seller, which characterizes the median market in our sample. In the market audits data, we don't find evidence markets with multiple sellers tend to feature firms that differentiate themselves on quality, which could reveal itself in the market audit through a bimodal distribution of "good" packets observed per shop (secret shoppers in our study observe two packets per shop). We also don't observe agrovets – commonly seen as more serious sellers of agricultural inputs – offering higher quality goods than general purpose retail shops.

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. For simplicity, we normalize the consumer's utility to $1-p$ if they obtain a high-quality good, and $-p$ if they obtain a low quality good. 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 can lead some to acquire this information and others to not acquire it.

Assume there is one local firm.[5] Consumers can buy from the local firm 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.[6] 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 obtain a high-quality good. An uninformed consumer believes they can obtain 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 before purchase. Uninformed consumers, on the other hand, cannot observe quality. They have beliefs about the average quality in the local market ($\hat{q}$) which we assume enters exogenously. Note that $\hat{q}$ could match the average quality that will be 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, 2023). In this model, beliefs that need not reflect rational

[5] We assume only up to one seller, which is clearly a simplification. However, the median market in our sample 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).

[6] We do not examine separating equilibria, as we do not find evidence that quality-verified seed sell for more, once key determinants of price are accounted for, including brand and county (Table D1).

expectations allows for inaccurate beliefs. Learning through repeated experience may especially be hindered when the difference in quality between low and high quality is not very large relative to the variance in observable outcomes. Thus, when the beliefs of uninformed buyers enters exogeneously we think of this as a model approximating short-run buyer and seller responses when beliefs about average quality change minimally in response to the seller's actual choice of quality.

We focus on the equilibrium in which the seller's optimal price causes: (1) uninformed buyers to buy locally, and (2) informed buyers to buy locally *only* if they observe a high-quality good.[7] Since informed buyers expect to get high quality from the outside option at cost $P_0 + S$, and uninformed buyers expect to get quality $\bar{q}$ in expectation at cost $P_0 + S$, so the following must both be true in this equilibrium:

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

We might default to thinking of the second expression as binding, in which buyers believe quality is better 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 in a similar study area in Kenya 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; in our setting, wholesale agents tend to have storefronts outside the local market in major towns. 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 observe high quality. They lose informed customers who are offered low quality as they opt for the outside option in this equilibrium. In this way the price that the local seller can offer is pinned down by the price offered in town, the search cost of going to town, and any beliefs that quality in the local market and quality in town differ (all taken as exogenous). In Appendix B for completeness we discuss the other possible equilibria and when they are dominated by this base case that we focus on.

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 – proportion q of the seller's stock is then high quality, while proportion $1 - q$ is low quality. The seller randomizes the quality of the product according to the choice of q, and they make a take-it-or-leave-it offer to each potential customer.

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

[7] See appendix for a discussion of other equilibria where informed buyers buy locally even if they are offered a low quality product, or where informed buyers use the outside option even if they are offered a high-quality product. These equilibria are less interesting to consider as a base case because receiving information does not change any buyers' behaviors and they do not match the setting where some buyers buy locally and some do not.

quality. The cost d may be thought of as potentially coming from several sources. For example, sourcing from a more reliable or formal seller or discarding packets that lack certain quality markers may result in higher per-unit costs or incur effort that is captured by d. $\frac{q}{1-\theta+q\theta}$ is the average quality among units that are sold if the seller chooses quality mix q, and informed buyers refuse if offered low quality (which happens proportion $q\theta$ of the time).

Let F represent the firm's fixed cost. We may think of F as also encompassing opportunity costs of entering the seeds business – for example, if the space, capital, and time required to engage in selling seeds crowds out other profitable business opportunities. We write x (the number of units sold) as a function of the firm's choice of price p and quality mix q to get the firm's problem:

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

The firm's solution 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, $q^* = 1$ is the firm's optimal choice of quality 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 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 strictly decreasing in the fraction of consumers that are informed as long as the firm offers less than the highest quality:

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

Anticipating this, a firm with $q^* = 0$ will rather not enter the market if expected profits π^* fall below zero:

$$(p-c)(1-\theta)N - F < 0$$

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

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

As θ rises, does the firm improve quality or exit? Putting together the inequalities above, we can see that the firm will improve quality rather than quit as θ rises if: $d < (p-c) - \frac{F}{N}$. That is, the upgrade cost must be "small enough" that the firm can be profitable selling only high-quality goods.

3.3 Summary of model implications

1. Buyers that become informed obtain higher quality goods. By the assumptions made, informed buyers can observe quality and can search until they find high quality, leaving the local market if necessary to find another seller.
2. Improved consumer information lowers local seller's expected profit ($\frac{\partial \pi}{\partial \theta} \leq 0$). The presence of more informed consumers reduces the local seller's market share as those detecting low quality leave to buy elsewhere.
3. Having more informed buyers does not affect market prices ($\frac{\partial p}{\partial \theta} = 0$). Sellers have local market power, but the presence of an outside option where consumers can also buy the product constrains the firm's ability to raise prices.
4. Uninformed consumers may also benefit from having more informed consumers in the local market. They benefit if the local seller is induced to upgrade quality rather than quit the market as the number of informed consumers increases. This happens if upgrade cost isn't "too high".

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, and (4) it is located in Bungoma, Busia, Kakamega, or Transnzoia counties in Kenya. To do so, we use a 2-stage sampling strategy–first randomly selecting sublocations within the study counties (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 in each sublocation.[8] This sampling strategy helps to minimize clusters of markets that are very close to one another, which would otherwise increase the likelihood of information spillovers between treatment and control sites, and it helps ensure that the sample of markets covers 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 selected market, all seed sellers and eight randomly chosen maize-farming households located 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. To select households, 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.

[8] In randomly selected sublocations in Transnzoia two markets were selected, due to logistical reasons and the relatively large size of sublocations in that county.

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 February 2020. We instead use the number of seed sellers in March 2021 to determine eligibility to be selected. For these sites we collected only data from market audit activities during the 2021 main planting season. No baseline surveys were conducted at these sites, and no follow-up seller surveys or household surveys were conducted at the pure control markets.

4.2 Randomized experiment

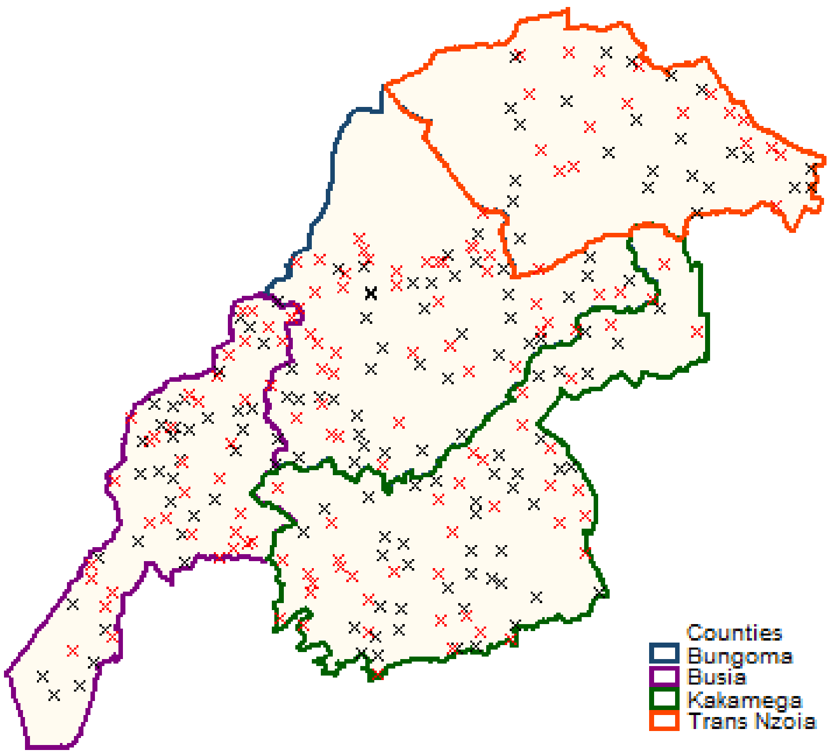
Sublocations were randomized into the following treatment groups, stratifying by county. We describe the treatment arms in greater detail below.

Treatment A	:	68 sublocations
Treatment B	:	68 sublocations
Control	:	146 sublocations
Pure Control	:	104 sublocations

Figure 1 shows the sampled markets within the four counties, with treated markets in red and control markets in black. The information campaign implemented at treated markets disseminates 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. These include the lot number and a sticker containing an SMS code that allows farmers to verify authenticity (see Appendix A for details). Using SMS verification allows farmers to detect at least two types of lower-quality seeds – uncertified seeds and old seeds that were unsold from previous seasons – and receive other information such as packaging date in a standardized format via SMS message. Hybrid seeds are typically purchased in closed packets produced by the manufacturer. Most of the quality markers associated with the information treatment can be observed for all brands of seed prior to purchase without opening the packet. In the Treatment B arm, community members additionally received some encouragement 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. While the intent was to encourage farmers to report issues to authorities with sanctioning power, in practice only two respondents indicated in follow-up surveys that they did so; research staff noted that many respondents may be hesitant to escalate to that level in cases of suspected uncertified seed. For this reason and to preserve statistical power the results that follow focus primarily on simple treatment versus control comparisons.

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

The model predictions from the previous section guide our thinking on the potential effects of the information treatment. All treated sites received information about using quality markers to detect quality-verified seed packets and reject packets missing one or more markers. This corresponds to an increase of θ in the model. 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 greater access to high-quality seed for even uninformed buyers.

Appendix A shows the (English version of) flyers which summarize the information that was communicated at treated market areas.[9] Those who received flyers were encouraged to pass extra copies along to family, neighbors, and friends to maximize the spread of the information within

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

the local community. Focus group discussions prior to project launch and conversations during implementation suggest that farmers perceived the provided information to be important for their livelihoods and that the information is accessible, particularly with the help of the flyer and family, friends, and neighbors. We note that one of the quality markers–the KEPHIS sticker–has multiple uses that overlap with the other markers and is more technically challenging to use properly; it requires a mobile phone and several steps to verify its authenticity and interpret the information encoded within the SMS response. 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 training.

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 a market begins with the seller offering $q = 0$. Only a large-enough change in θ that causes a certain threshold to be crossed will induce sellers to change quality or entry decisions. In practice, though, the data shows little evidence of such threshold effects from the variation in treatment intensity. Therefore, the following section 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 reached approximately 19.7% of local seed customers (see Table D2). [10] 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, on average about 49% of the local (within 1km) adult population received a flyer, and 22% were treated directly face-to-face.

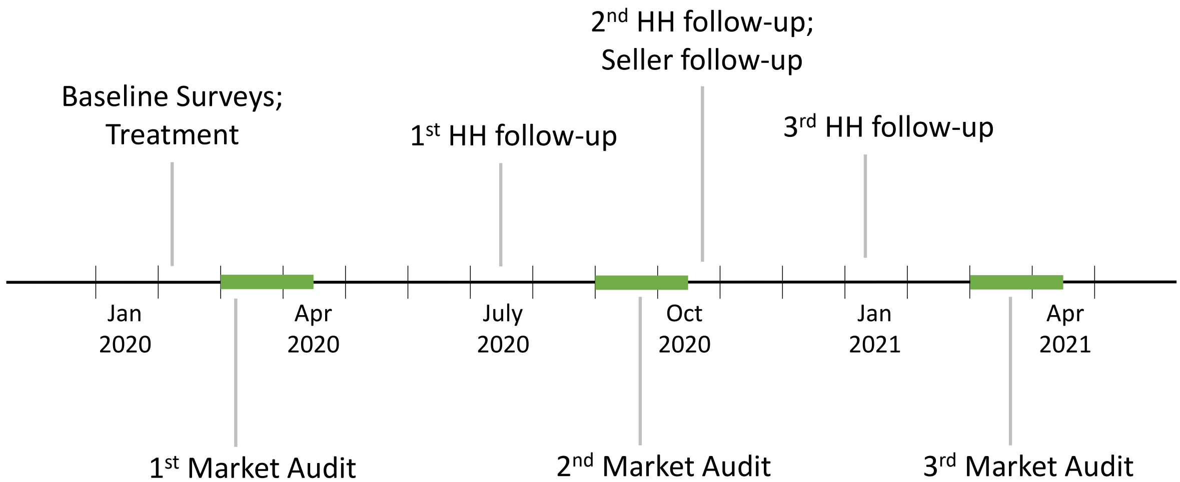
4.3 Data Collection and Timeline

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 contacted village elders, baseline household and sellers surveys were carried out. The information treatment was typically delivered later in the same day, with research staff speaking at a convening in the mid-afternoon in a public location and door-to-door dissemination of information proceeding afterwards. Market areas that are assigned to receive additional visits for information dissemination would then have a team member return for additional door-to-door dissemination activities on subsequent days.

To measure seed quality and possible adjustments in quality, price, and entry by sellers, the team 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

[10] We use several pieces of data to estimate this figure, including: number of sellers in a market, number of customers per seller from the previous year's main planting season, and percentage of local customers who shop locally for seeds.

Figure 2: Timeline



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

markets, view up to two packets, and purchase one packet. They used realistic scripts to purchase "seeds 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. Based on the field staff's experience implementing these activities, we do not expect that secret shoppers were 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. Due to budgetary limitations, seeds were not subjected to DNA testing or to field trials.

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 present but 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 displayed in the SMS response 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.

In the fourteen months following the information treatment, the team completed three rounds of household follow-up surveys via phone to document household seed purchasing choices and agricultural outcomes. Households self-reported harvested amounts for each maize variety they planted, and these were used to compute maize yields. Survey timing was set strategically around key dates in the typical crop timeline 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). During each survey round, each household was called several times if there was no response. Data on maize harvest was collected from 85% of respondents, and we do not find indications of differential attrition between treatment and control households (Table D3). To guard against concerns that self-reported yields may be inaccurately reported, we used self-reported yields to inform our sample size, anticipating noise from mismeasurement.[11] 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.

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

During follow-up surveys and market audit activities, enumerators were not made aware of the treatment status for the sites and respondents that were targeted by these activities. Lab testers were also not made aware of the seed sources – all seed samples were repackaged into plain paper bags labeled with unique alpha-numerical identifiers that were used internally by the research team and do not reveal any seed or site characteristics.

4.4 Baseline Balance, Spillovers, and Other Threats To Identification

We test for baseline balance among market areas in our sample and find that household heads in treated sublocations were slightly less likely to have completed primary or secondary school (Table D4). However, when jointly testing 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=0.51). To address the possible concern that imbalance in education levels could skew the experimental results, the household level results below showing treatment effects will control for primary and secondary education of the household head, though results are similar when not doing so.

In Appendix C, we explore the possibility that the experimental design may have been contaminated in two ways. First, we ask if baseline activities could have influenced sellers by altering their beliefs about the likelihood their products will be scrutinized. Second, we ask if information spillovers could have affected buyer and seller knowledge and behavior in control sites that are near treated sites. We do not find evidence consistent with these two hypotheses and so in the main empirical results shown in section 5 we focus on comparisons between treatment and control sites.

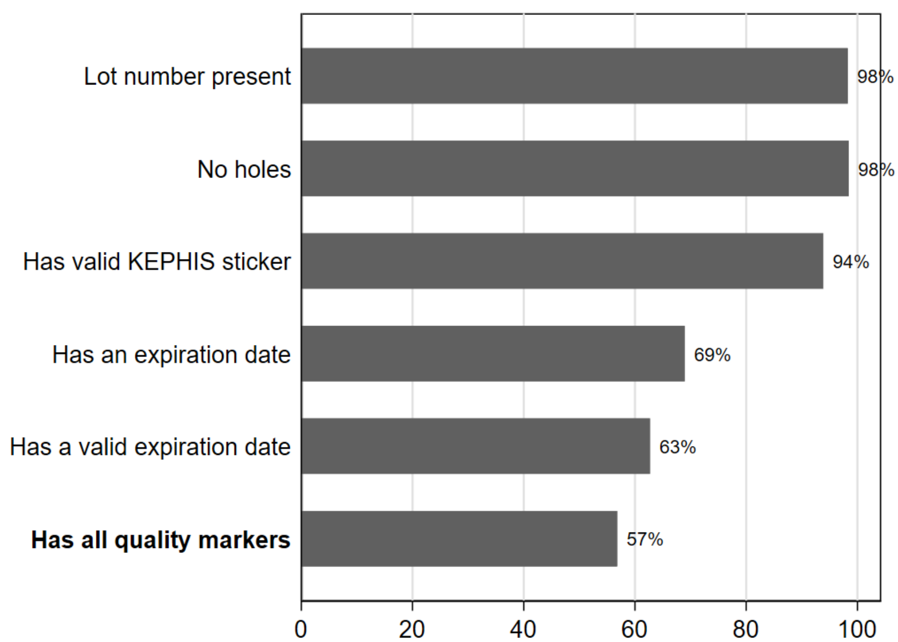
5 Results

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 show that farmers are able to adjust their purchasing behavior in response to the information treatment, consistent with channels in the model. We also examine agricultural outcomes, finding evidence that informed farmers improved maize yields, particularly in more remote markets (where substandard quality was a bigger problem at baseline), and among more educated farmers (who retained the information better). These findings correspond to model prediction #1 (see section 3). In subsection 5.3, we examine effects on sellers' decisions. We document a sizable effect of treatment on seller exit, consistent with the negative effects of informed buyers on profit from model prediction #2. We do not find effects on seed prices or on seed quality, which correspond to model predictions #3 and #4 in a world where upgrade costs are high.

5.1 Summary statistics: quality markers and their correlates

Figure 3 presents summary statistics for quality markers, observed during the market audit exercise. These descriptive statistics provide 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 or refuse. Overall, 57% of packets that our team observed had all quality markers, while 43% were confirmed to be missing one or more. Most of the missing quality markers were due to not having an expiration date or having an invalid expiration date. Missing or invalid KEPHIS stickers and the other observable markers also contributed.

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. The sample includes 1508 seed packets purchased during market audit activities in the main season in 2020, the short season in 2020, and the main season in 2021. It includes observations representing 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 57% of packets had all quality markers present.

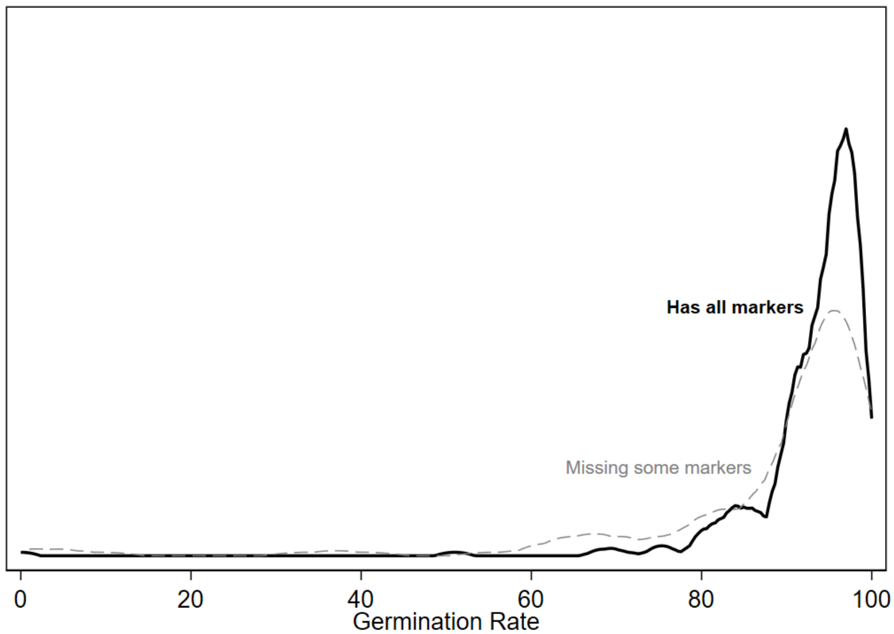
We next ask if the presence of quality markers correlates with aspects of seed quality measured in lab tests. Lab tests are not subject to self-reporting biases and testing staff are blinded to the seed source, variety, and any characteristics of the packaging. Table 1 shows that observable quality markers are uncorrelated with purity, but strongly correlated with germination rate. Nearly all contents of seed packets consisted of whole unbroken seed, regardless of the presence of all quality markers. However, packets with all markers germinated better. This is illustrated in Figure 4, which shows the distributions of germination rate for packets with and without all quality markers. Packets with all quality markers are substantially more likely to have a germination rate that is greater than 90% as was required for all production lots during certification. To examine which quality markers drive this correlation, we break down the analyses using each quality marker separately (Table D5 and Table D6). This shows that the the relationship between quality markers and germination rate is driven primarily by the presence of a valid expiration date. We view the observed 5.2 percentage point (or 5.9%) overall difference in lab tested germination rates as a lower bound for the changes in yield that could be achieved should farmers switch from seeds that are not quality-verified to quality-verified seeds. Note that seeds that are not quality verified can result in lower yields without lower germination rate at all (e.g. well-stored non-hybrid seeds may be expected to behave in this way). Lower germination rates may also not account for other characteristics of seeds that affect yield, which is expected to correlated positively with germination rate and whose effects on yield can far exceed the differences in germination rate (Ghassemi-Golezani and Mamnabi, 2019). For example, expired seeds could exhibit lower germination rates but also feature nutrient degradation, which could result in a seed that germinates but has poor initial root growth and subsequent plant development. The relatively modest 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 had extremely low germination rates, this would be salient to farmers, who then may be able to more easily learn about inferior seeds and adjust purchasing behavior in subsequent planting seasons. Some farmers experiencing non-germinating seeds may replant to fill in parts of their fields after observing where germination failed. This may partially reduce losses due to poor germination but is an imperfect solution because replanting may be misaligned with the rains.

While we cannot be certain of the specific causes for substandard quality, the observed issues with expiration dates and the germination test results would at least suggest that the presence of expired seeds is a likely issue. Expired seeds would have had more time to be exposed to improper temperature and moisture environments, and this would explain lab test results on germination rates and match evidence from Uganda suggesting that actors downstream in the supply chain–as opposed to manufacturers–are likely responsible for most of the substandard seeds in the market (Barriga and Fiala, 2020).

The relationship between quality markers and lab-tested germination rate differs by location. Table D7 shows that more remote market areas (greater than the median distance from the county capital) see germination rates that are lower overall and farther from national standards for minimum germination rate. At the same time, in these areas quality markers are more informative to farmers: having all quality markers is associated with a greater gain in germination rates in more remote markets.

These correlations between observable markers and lab-tested germination rates do not capture

Figure 4: Distribution of germination rates



This figure shows the distributions of germination rate for (1) seeds from packets featuring all quality markers, versus (2) seeds from packets that are missing one or more quality marker. The sample includes 467 seed packets that were purchased by secret shoppers from control sites and tested in a lab by KEPHIS staff. The sample includes packets purchased during the main season in 2020, the short season in 2020, and the main season in 2021.

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 should in theory stand to gain in agricultural output if they can successfully apply a strategy that uses these observables to refuse lower quality seeds in favor of better quality-verified seeds.

5.2 Effects on household knowledge and seed purchases

The following sections show treatment effects on household, seller, and market-level outcomes 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 respondent i, during season s, in county c; *Treated* is a dummy that equals 1 if the sublocation of the respondent was assigned to the treatment group; X_{isc} is a vector of controls, which is described in the pre-analysis plan. For households this includes the household head's gender and age, county fixed effects, and the value of the dependent variable at baseline. In the specifications shown below, we control also for planting-season fixed effects, and dummy variables for primary and secondary school completion by the household head

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 control sites and tested in a lab by KEPHIS staff. The sample includes packets purchased during the main season in 2020, the short season in 2020, and the main 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). The independent variable equals 1 if all quality markers are present and 0 otherwise. Table D5 and Table D6 show similar results when considering each quality markers individually.

as discussed in subsection 4.4. Sublocations are of varying sizes, especially in different counties, and so not all markets had the same probability of being selected into the sample in our two-stage sampling procedure. Observations in the analysis 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.

We first examine treatment effects on several measures of household knowledge and usage of quality markers to verify seed quality, as shown in Table 2. Rather than asking direct questions about specific quality markers, enumerators asked respondents an open-ended question ("During the main growing season, were you able to verify the quality of the seeds?") with follow-up questions. This unprompted approach has the benefit of being unlikely to expose respondents in the control group to the information from treatment. However, compared to a prompted recall approach, these data may under-report respondent knowledge, a pattern for respondent recall that has been found in a variety of other contexts (Romaniuk, 2006; Waller et al., 2004). Certain aspects of the quality of respondents' knowledge was left unprobed as a result.

Treatment substantially increased knowledge about visible quality markers, being associated with increased recall of specific elements by 32% to 108%. Household-level effects appear to be heterogeneous. For households with more highly educated household heads with primary school or secondary schooling completed, knowledge (by all measures) increased substantially more due to treatment (Table D8). These increases in knowledge reflect relatively low levels of consumer knowledge about quality markers at baseline. At the time of the study, KEPHIS had yet to run a major campaign to inform buyers of the verification system while the stickers themselves say little about their purpose and expiration dates feature highly inconsistent formatting (see Appendix A).

While demonstrating the efficacy of the information treatment, this result is subject to a few of caveats that affect its interpretation. First, these knowledge measures and self-reported usage of quality markers were collected via a phone interview taking place about five to six months after the information treatment. These data on household knowledge were collected in July and August 2020, and respondents did not know in advance when they would be called by the research team. 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. While in-person activities 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 several months after planting.

Second, the team's qualitative observations from piloting and implementing the information treatment strongly suggested that, in practice, the quality of knowledge–beyond simply identifying a factor included in the training–also mattered for whether a respondent would be prepared to apply the knowledge to purchasing decisions. For example, merely mentioning expiration date as relevant may not mean that the respondent is able to find an expiration date that is embedded within a lot number or an e-verification SMS message, or infer an expiration date when only a packaging date is available. Somewhat relatedly, the possession of knowledge about quality markers may not translate one-for-one to the expected behaviors. For example, a respondent who recognizes the quality markers nevertheless may not ascribe very much weight to them, and consequently the presence (or lack) of quality markets may not greatly influence the respondent's purchasing decisions.

Third, the relevant information set used when purchasing seeds may not be wholly captured by the respondent's knowledge. For example, if the respondent enlists the help of other people in the seed purchasing decision, or if they delegate seed purchasing to another family member or a friend, then the household's seed purchase may reflect information that the respondent does not personally possess.

We also examine the possibility that households at control sites near treated sites may have experienced gains in knowledge due to spillovers. As shown in Appendix C, we do not find evidence of spillover effects within 2km, 4km, or 6km, suggesting that any spillover effects that may have been present are too small to be detectable. Partly by design in using a multi-stage sampling strategy, the sample has very few markets that are clustered closely together (e.g. within 2km of one another). This means that few markets are at distances where we might be most concerned that spillovers could affect estimates of treatment effects. At the same time, it 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 spillovers not detected due to limited statistical power affect control sites similarly but to a smaller degree than the treatment sites, one could think of the estimated treatment effects as being lower bounds.

After the main 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

Table 2: Respondent reported usage and knowledge of quality markers

	(1) Able to verify	(2) Expiration mentioned	(3) E-verification mentioned
Treated	0.192***	0.048**	0.242***
	(0.026)	(0.022)	(0.028)
Observations	2018	2002	2002
Control Mean	0.18	0.15	0.29
Treatment Effect (%)	108.77	32.02	84.31

The table shows the effects of information treatment on self-reported usage of markers and respondent recall of knowledge included in the information treatment. The sample includes respondents that planted hybrid maize in the main growing season of 2020. The dependent variables are (1) whether the respondent indicated they were able to verify seed quality for the main planting season in 2020 (=1 if yes), (2) whether the respondent mentioned expiration date as a way to verify seed quality (=1 if yes), and (3) whether the respondent mentioned e-verification as a way to verify seed quality (=1 if yes). Measures of knowledge and ability to verify quality were collected via phone survey in July to August 2020, which was approximately 5 to 6 months after the information treatment was implemented. Standard errors are clustered by sublocation. Results broken down by the household head's level of education are shown in Table D8.

seeds they purchased or where they purchased them.[12] While suggestive, these responses indicate that many respondents perceived that the information provided affected their purchasing decisions.

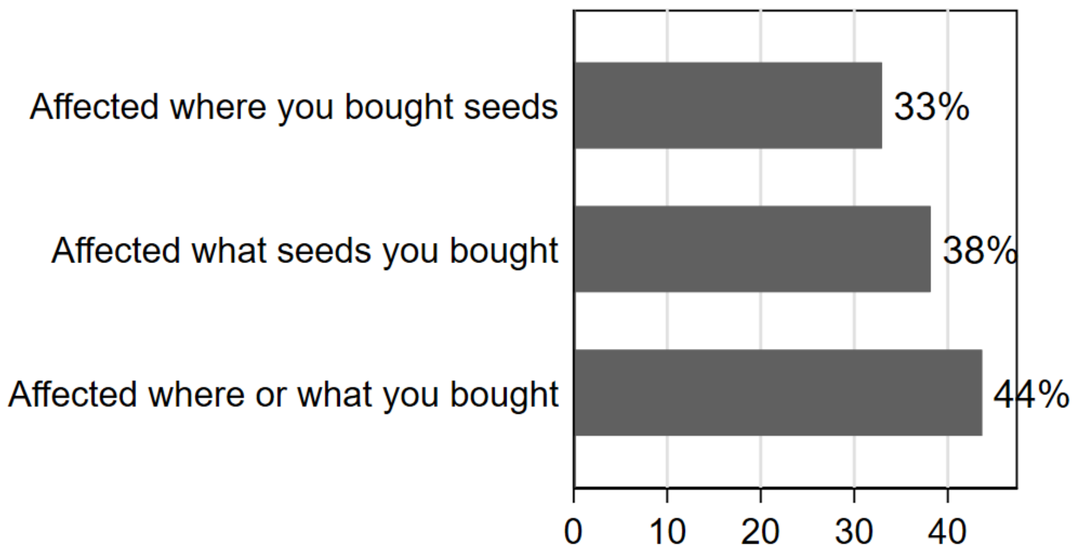
Table 3 shows results examining whether treatment affected whether and where households bought hybrid maize seed. Columns 1 and 2 show that households shifted their source for seeds, with a 16% reduction in the likelihood the household sources seeds from the local market; households substitute by being more likely to purchase seeds outside the local market. Gharib et al. (2021) provides complementary results that suggest farmers have higher willingness to pay for seed packets that have certain quality markers present.[13] 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. The treatment protocol emphasized the importance of looking for quality markers before purchase, but it did not comment on local sellers versus sellers in other markets or any other buying strategy. Overall, households were not more likely to buy hybrid seeds, as seen in column 3 of Table 3. 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 other market imperfections (such as credit constraints or incomplete insurance markets) may act as important barriers that limit adjustments on the adoption margin (Karlan et al., 2014).

We next examine how the information campaign affected household agricultural outcomes. To-

[12] Based on the open-ended portion of the response, it appears that some farmers interpreted the questions such that the answer would be "no" if they did not adjust their choice of seller or variety, even if in theory checking for quality markers and finding them absent would hypothetically have caused them to change where they purchased seeds or what packet they purchased.

[13] 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.

Figure 5: Do treated households think the information helped?



The following questions were asked to 742 respondents that planted maize in treated market areas during the main season of 2020: (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 or both.

tal harvested amount and yield (kilograms harvested per acre) are the main outcomes we consider. 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 4 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. Treatment effects are concentrated in more remote areas, as illustrated in Figure D2, which shows results of four pre-specified analyses of treatment effect heterogeneity. In more remote areas, we note that baseline seed quality was worse and quality markers are associated with greater quality gains (Table D7). Figure D2 also shows somewhat higher gains for households with a more educated household head. This is consistent with the data showing that more educated respondents appeared better able to retain the information contained in the training. 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

Table 3: 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
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. Specifications control for gender, age, and education of household head, dependent variable measured in the 2019 main season, and county and planting season fixed effects. Robust standard errors clustered by sublocation are shown in parentheses.

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.

We do not see treatment effects on the prices that households paid (Table D9). This suggests that the estimated agricultural gains represent welfare gains for the households. An important caveat, though, 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 (Table D15). 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 D11) and when using a post-double selection LASSO approach to selecting control variables to reduce noise in the estimates (Belloni et al., 2013) (Table D12).[14] As described in the pre-analysis plan, we also examine results when dropping households that may be thought to be less likely to switch sellers – namely, households with a close personal relation with their usual seed seller (Table D13), and households that previously depended on credit from the seller to purchase seeds (Table D14) – and we find similar results.

To check the implications of these estimated average 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%.

[14]The procedure selects controls among those discussed in subsection 5.2, baseline characteristics shown in Table D4, and two and three way interactions among these variables.

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.[15] 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 with what might be expected (given evidence from poor quality seeds in Uganda, or evidence on the performance of old seeds from controlled trials).

Treatment effects on farmers' self-reported germination rates do not show significant effects (Table D10), though we note that the vast majority of respondents estimate germination rates in multiples of five percentage points, pointing to the relative inaccuracy of this measure of germination rates. This could be because improvements in germination rates were actually relatively small, as would be expected given the modest differences in lab-tested germination rates between quality-verified packets and packets that are not quality-verified. If these small declines in germination rates are due to poor storage, this would still be expected to generate large differences in yields (Ghassemi-Golezani and Mamnabi, 2019). It is also possible that packets that are not quality-verified also have higher rates of non-hybrid seed (e.g. as was documented in Ethiopia in Michuda et al. (2022)), which could germinate as well as genuine hybrid seed while exhibiting lower yields at harvest.

Table 4: Household agricultural outcomes

	(1) Kgs harvested	(2) Yield
Treated	57.75**	54.84**
	(29.05)	(27.39)
Observations	3807	2443
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. Specifications control for gender, age, and education of household head, dependent variable measured in the 2019 main season, and county and planting season fixed effects. Robust standard errors clustered by sublocation are shown in parentheses.

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

5.3 Effects on seller exit, quality choice, and prices

We next examine the effect of the information treatment on seller decisions, including decisions for entry and exit, setting quality, and setting prices.[16] As shown in Table 5, the number of sellers per market decreased by about 0.27 sellers in the SGS2020 and LGS2021 seasons due to treatment. This is about a 17% percent decrease in the number of sellers. Figure D1 shows treatment effects on seller entry and exit by planting season, showing that sellers did not respond in the first season after treatment as buyers did, but rather began adjusting one planting season later, perhaps in response to observing changes in purchasing behavior in the first season. The pattern of sellers exiting due to treatment is consistent with model prediction #2, where informed buyers cause sellers to expect increased costs if they opt to upgrade quality and fewer customers if they do not; this 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 (Table D16). Firm exit was driven disproportionately by firms that reported below median profit at baseline (Table D17).

Table 5: Seller entry/exit

	Number of sellers at market
Treated	-0.270*
	(0.138)
Observations	495
Control Mean	1.58
Treatment Effect (%)	-17.07

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 and offering hybrid maize seeds for sale, as observed by secret shoppers. The specification controls for county and planting season fixed effects. Standard errors are clustered by sublocation. Table D17 shows related results in a sample of surveyed sellers, separated by level of baseline profit. Figure D1 breaks down results by planting season.

Consistent with the model predictions together with the firm exit results, the data show no detectable effects on visible quality markers or on purity and germination rates of sampled seeds obtained by covert shoppers. As Table 6 shows, we see about a one percentage point increase in packets with all quality markers present, though not statistically significantly different from zero. Treatment did not affect seed purity or germination rate of seeds offered in the local market. We also do not see evidence that treatment affected price levels or price dispersion as revealed by the

[16] 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 limited time in March 2020 available before field operations were forced to stop, the field team visited a subset of 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. The short season in 2020 and the main season 2021 are unaffected by these data limitations.

secret shopper data (Table D18).

Table 6: Quality markers and lab tests

	(1) Has all quality markers	(2) Purity	(3) Germination Rate
Treated	0.012	-0.277	-0.539
	(0.028)	(0.274)	(0.753)
Observations	1212	878	878
Control Mean	0.57	99.85	91.84
Treatment Effect (%)	2.16	-0.28	-0.59

This table shows treatment effects on quality of seeds offered to secret shoppers that posed as uninformed buyers. The sample includes markets in the main season in 2020, the short season in 2020, and the main season in 2021. 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). Specifications control for county and planting season fixed effects. All standard errors are clustered by sublocation.

These results indicate that improved consumer information did not induce sellers to upgrade quality, and as a result uninformed buyers in treated markets did not enjoy better access to higher quality seeds through market mechanisms. As the framework in section 3 suggests, this is the expected result if the cost of upgrading quality is too high. In the next section we discuss some reasons that may explain this and discuss appropriate policy responses.

6 Policy Implications

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 least partly by research needs, such as simplicity of the intervention, targeted exposure to information to limit spillovers, and bundling treatment with data collection 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. Assuming no information or market-based spillovers to people not directly trained, a 1% increase in productivity would be sufficient to justify implementing this information campaign.

Of course, we take such cost-effectiveness figures with a grain of salt. Firstly, point estimates come with confidence intervals, and extrapolating will depend on being able to effectively target, especially given the heterogeneous effects seen even within our study sample. Secondly, the cost estimates that are discussed above completely omit the costs of establishing the pre-existing regulatory structure that makes the information treatment possible. In many ways, the information treatment is a form of "last mile" outreach to end-users, building on top of extensive government efforts to certify seeds nationwide. Lastly, we would expect that other related information interventions may have similar effects and may scale better. For example, updating the messaging in the 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 more cost-effectively disseminate information to consumers.

If we view the findings through the lens of the simple model of section 3, a key policy goal would be to lower the costs to firms of upgrading quality. When it comes to improving overall access to high quality inputs, this would be complementary to improving consumer information. A sufficiently large increase in the number of informed consumers could unlock large positive spillovers to uninformed consumers if they can induce the local seller to upgrade quality. On the other hand, if sellers are induced to exit, then this simply removes an option for uninformed buyers to choose from and could be welfare-decreasing for them. A key question is – why are sellers behaving as if upgrade costs are high? The appropriate policy response depends on the answer to this question.

One possibility could be that sellers are not very knowledgeable about seed quality markers, and so they have difficulty sourcing high quality seeds. The effort required to learn how to screen based on quality markers could be viewed as a cost to upgrading quality. In this case, training or other information dissemination activities targeted to sellers could be helpful. We note that to the extent that sellers did not have at baseline the kind of knowledge disseminated in the information treatment, it is highly likely that the information treatment trained them to use the quality markers. Villages in the study area are not particularly large, and the main group training session was done at each treated market near the market center, where seed sellers also tended to be located. We also don't see that agrovets–which are viewed as more specialized sellers of agricultural products–tend to offer products of higher quality. In our view, these observations make it less likely that lack of seller knowledge about quality markets could fully explain the pattern of results.

Another possibility is that monetary costs of upgrading quality are high and could explain seller exit. Retail sellers have a choice to source seeds from authorized agents (official agents of the seed manufacturer), unauthorized agents (for example, acquiring previously unsold seeds from a third party informal distributor), or to save unsold seeds from previous seasons. If authorized agents offer higher quality but also sell for higher prices, then this could be one source of high upgrade costs. Furthermore, anecdotal evidence from research team visits to wholesalers in town suggest that even authorized upstream agents sell packets that are imperfect and are missing quality markers. This would suggest that stocking more quality-verified packets may require retailers to identify and discard non-compliant packets, which would result in higher per-unit costs. Relatedly, it could be that low-quality packets can be sourced at very low marginal cost such that the upgrade cost is quite high for that reason. Either way, these explanations would suggest that interventions within the supply chain that reduce the prevalence of unsold seeds or improve quality offered by wholesalers could be helpful.

One more possibility that we cannot examine directly is that firms may have substantial opportunity costs of stocking seeds. If these opportunity costs are substantial, then even small upgrade costs could lead to firm exist as a profit-maximizing response. If for example, the source of high opportunity costs arises from alternative business opportunities in the face of limited credit access, then policies that address limited access to credit could help improve the functioning of markets for hybrid seeds too.

7 Conclusion

In this paper, we study empirically the effects of improving consumer information about product quality in the hybrid maize seed market in rural Kenya. We evaluate a randomized market-level information campaign to quantify effects on both seller and buyer behavior. We monitored seller entry and exit, product quality, and prices for over one year to allow for sellers to respond over time to the increase in the number of informed buyers in the local market. First, we show 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 production. Second, improved information caused sellers to exit the market, and we do not observe effects on prices or quality among the stayers. This is consistent with a model in which the cost to the firm of upgrading products from low to high quality is substantial, and firms prefer to quit the market rather than be induced to offer high quality.

Our findings show that policies that help improve consumer information may be beneficial to supporting the productivity of small-holder farmers. For example, reforming existing requirements for information displayed on product packaging for consumers to promote greater understanding about certifications may be helpful. More generally, the findings inform how consumer information can play a role in enforcing regulations in an environment with poor enforcement of product standards. The absence of quality upgrading on the part of sellers in this paper's setting points to remaining challenges to overcome the difficult learning and reputation-building environment. Efforts to reduce barriers to firms to upgrading quality to meet standards may be especially beneficial and could complement efforts to improve consumer information.

A Appendix A: Information campaign details

Figure A1: Information Treatment 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: Information Treatment Flyer: Supplement for Treatment B

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

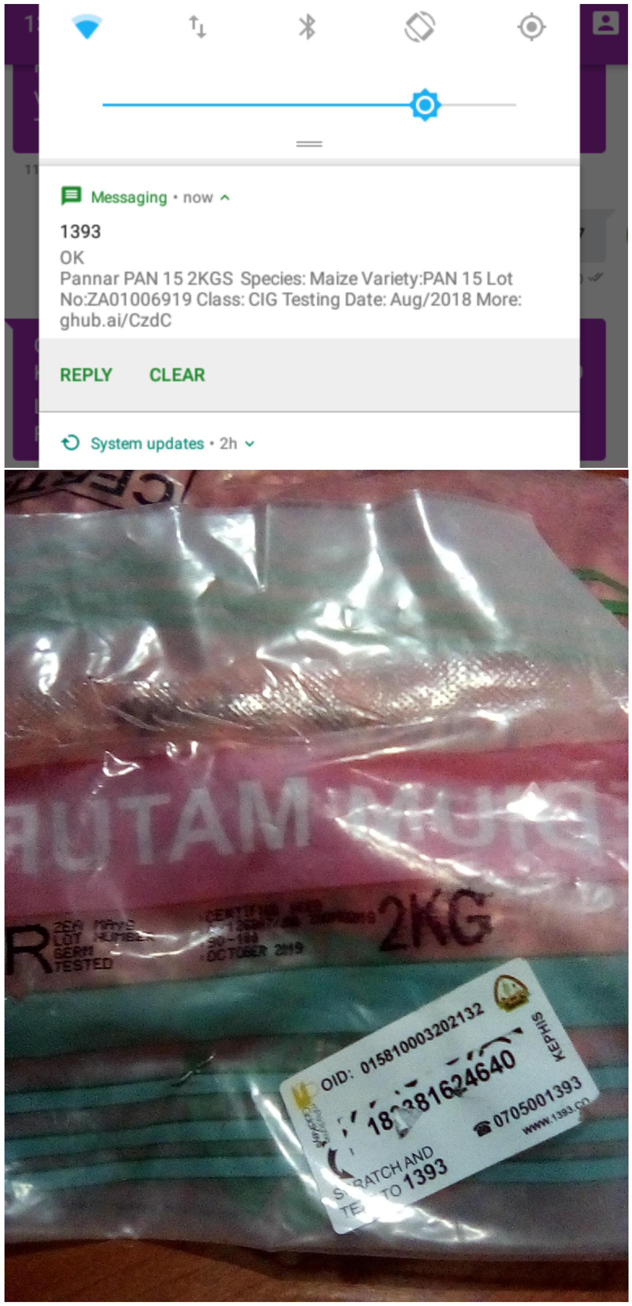
- Kenya Plant Health Inspectorate Service (KEPHIS)
  Phone Number: 020-3597209
- Anti-Counterfeit Authority (ACA)
  Phone Number: 020 2280111
- Your assistant chief

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: Example: Seed packets with (left) and without (right) quality marks

Figure A4: Example: Using 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 equilibrium responses by consumers 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 the derivative with respect to q, we get:

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

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

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

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 a firm whose optimal choice of quality is zero will choose to exit the market if $\theta > 1 - \frac{F}{(p-c)N}$. 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 offering low quality will opt to upgrade quality over quitting as θ rises if:

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

That is, the firm will upgrade quality if there are enough informed consumers, and if the cost of upgrading is not too high. This inequality is critical for the spillover effects of improved consumer information. If the upgrade cost is low enough, then informing some of the consumers in the market will cause all the consumers in the market to access high quality seeds. In this toy model, there are three different types of consumers that may experience changes in utility.

Case 1: As θ increases, it induces the firm to switch from $q = 0$ to $q = 1$. Here, improved consumer information benefits all consumers.

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

Case 2: As θ increases, it induces the firm to exit. Here, the always informed do not benefit from improved consumer information, and the never-informed are not better off if their choice of seed is informed by rational beliefs about average quality in the market). [17]

- The always-informed have no change in utility, as they left the local market to purchase seeds even before the increase in θ.
- The newly-informed benefit, going from getting $-p^*$ utility to $1 - P_0 - S$, as they switch to buying high quality seeds from the outside option
- The never-informed may not benefit, going from getting $-p^*$ utility to $\bar{q} - P_0 - S$ utility.

B.2 Conditions for equilibrium

In section 3 and in the previous section, we restricted ourselves to examining an equilibrium in which sellers set price low enough to entice uninformed buyers and informed buyers who observe high quality to buy locally, but not informed buyers who observe low quality. Below, for completeness we briefly discuss alternative cases where the seller sets price higher or lower than in the base case, and conditions under which the base case would represent the seller's best pricing strategy.

Alternative case #1: the firm chooses a lower price to induce even informed consumers who observe low quality to buy locally. Those consumers get $1 - P_0 - S$ in the outside option and $-p$ when buying from the local seller. Therefore, this informed consumer who gets low quality will stay if: $-p \geq 1 - P_0 - S$, or $p \leq P_0 + S - 1$. Note that if this holds, then uninformed buyers will also buy locally, since they do so if $\hat{q} - p > \bar{q} - P_0 - S$ and this is guaranteed. In this case, the firm gets

[17] Another way to think of this is that it is possible uninformed consumers could benefit in this scenario, but only if their beliefs about quality are incorrect such that the removal of the local low-quality option serves to remove a tempting but objectively worse option.

all potential customers, even if q is set at 0. So $q^* = 0$. The firm maximizes: $(p - c)N - F$ such that $p < P_0 + S - 1$. The solution is to pick the highest price that satisfies the constraint, with $p^* = P_0 + S - 1$. Firm profit is: $(P_0 + S - 1)N - F$.

Alternative case #2: The firm chooses a price high enough that even informed buyers that receive high quality opt to take the outside option. This happens if $1 - p < 1 - P_0 - S$. Note that if this is true then $\hat{q} - p < \bar{q} - P_0 - S$ holds as well, as long as $\bar{q} > \hat{q}$ (i.e. uninformed buyers believe that they are more likely to receive a high quality product in town). In this scenario, with no customers, the firm exits and receives zero profit.

Alternative case #3: The firm sets a price that is high enough to deter uninformed buyers from buying locally ($\hat{q} - p < \bar{q} - P_0 - S$), but low enough that informed buyers still buy locally when they observe high quality ($1 - p \geq 1 - P_0 - S$). This case is possible when $\bar{q} > \hat{q}$. In this case, the firm maximizes $(p - c - d)\theta N$. Only high quality seeds get sold, so the firm might as well choose $q = 1$. The price needs to be low enough so that informed buyers receiving high quality stay: $1 - p \geq 1 - P_0 - S \implies p < P_0 + S$. We also need price high enough that uninformed buyers will leave the local market: $\hat{q} - p < \bar{q} - P_0 - S \implies p > \hat{q} - \bar{q} + P_0 + S$. So price must be set with $\hat{q} - \bar{q} + P_0 + S < p < P_0 + S$. Profit is maximized at the highest price within this range with $p^* = P_0 + S$. Firm profit is $(P_0 + S - c - d)\theta N - F$.

If we compare the firm's profit in the base case and in each of these alternative equilibria, we see that firm profit varies:

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

Alternative case #1: $(P_0 + S - 1)N - F$

Alternative case #2: 0

Alternative case #3: $(P_0 + S - c - d)\theta N - F$

For the discussion in section 3, we must assume that the parameters $P_0, S, \hat{q}, \bar{q}c, d$ are such that profit under the base case is higher than in alternative case #1 and #3.

C Appendix C: Spillovers Effects And Effects of Baseline Surveys

In this section, we first 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 on the right-hand side. We find little evidence that the information treatment had spillovers on household knowledge (Table C1), on firm entry and exit (Table C2), or on seed quality (Table C3).

Table C1: Respondent reported usage and knowledge of quality markers (Spillovers)

Panel A: Spillovers within 2km

	(1) Able to verify	(2) Expiration mentioned	(3) E-verification mentioned
Number of treatment sites within 2km	0.002	-0.016	0.037
	(0.016)	(0.027)	(0.034)
Number of control sites within 2km	0.010	-0.006	0.062
	(0.017)	(0.029)	(0.038)
Observations	2425	1224	1224
County FE	Yes	Yes	Yes
HH Baseline Controls	Yes	Yes	Yes

Panel B: Spillovers within 4km

	(1) Able to verify	(2) Expiration mentioned	(3) E-verification mentioned
Number of treatment sites within 4km	-0.007	-0.029*	0.007
	(0.009)	(0.015)	(0.019)
Number of control sites within 4km	0.016*	0.015	0.020
	(0.008)	(0.013)	(0.016)
Observations	2425	1224	1224
County FE	Yes	Yes	Yes
HH Baseline Controls	Yes	Yes	Yes

Panel C: Spillovers within 6km

	(1) Able to verify	(2) Expiration mentioned	(3) E-verification mentioned
Number of treatment sites within 6km	0.005	-0.005	0.015
	(0.006)	(0.012)	(0.012)
Number of control sites within 6km	0.014***	0.014	0.019**
	(0.005)	(0.009)	(0.010)
Observations	2425	1224	1224
County FE	Yes	Yes	Yes
HH Baseline Controls	Yes	Yes	Yes

The table shows estimates of spillovers effects of the treatment on self-reported usage of markers and respondent recall of knowledge included in the information treatment. The sample includes respondents that planted hybrid maize in the main growing season of 2020. The dependent variables are (1) whether the respondent indicated they were able to verify seed quality for the main planting season in 2020 (=1 if yes), (2) whether the respondent discussed expiration date as a way to verify seed quality (=1 if yes), and (3) whether the respondent discussed e-verification as a way to verify seed quality (=1 if yes). Measures of knowledge and ability to verify quality were collected via phone survey in July to August 2020, which was approximately 5 to 6 months after the information treatment was implemented. Standard errors are clustered by sublocation.

We next examine the possibility that baseline surveys themselves influenced market participants, even at the control sites. To do so, 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). The market audit data for the main planting season in 2021 includes data for both control markets and pure control markets. We find no statistically significant differences between control and pure control markets in the numbers of sellers (Table C4), seed quality (Table C5), or market prices (Table C6).

Table C2: Seller entry/exit (spillovers)

Panel A: Spillovers within 2km

	(1)
	Number of sellers at market
Number of treatment sites within 2km	-0.271
	(0.194)
Number of control sites within 2km	0.067
	(0.257)
Observations	257
County FE	Yes

Panel B: Spillovers within 4km

	(1)
	Number of sellers at market
Number of treatment sites within 4km	-0.111
	(0.160)
Number of control sites within 4km	-0.076
	(0.099)
Observations	257
County FE	Yes

Panel C: Spillovers within 6km

	(1)
	Number of sellers at market
Number of treatment sites within 6km	-0.071
	(0.082)
Number of control sites within 6km	-0.053
	(0.074)
Observations	257
County FE	Yes

This table estimates spillover effects from treatment on seller entry and exit. 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 and offering hybrid maize seeds for sale, as observed by secret shoppers. Specifications control for county and planting season fixed effects. All standard errors are clustered by sublocation.

Table C3: Quality markers and lab tests (spillovers)

Panel A: Spillovers within 2km

	(1) Has all quality markers	(2) Purity	(3) Germination 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

Panel B: Spillovers within 4km

	(1) Has all quality markers	(2) Purity	(3) Germination 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

Panel C: Spillovers within 6km

	(1) Has all quality markers	(2) Purity	(3) Germination 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

This table estimates spillover effects on quality of seeds offered to secret shoppers that posed as uninformed buyers. The sample includes markets in the main season in 2020, the short season in 2020, and the main season in 2021. 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). Specifications control for county and planting season fixed effects. All standard errors are clustered by sublocation.

Table C4: Firm entry/exit: control vs pure control

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

This table examines the influence of baseline survey activities on seller entry and exit. The sample includes control and pure control markets in the main season in 2021. The dependent variable equals the number of sellers present at that market and offering hybrid maize seeds for sale, as observed by secret shoppers. Specifications control for county and planting season fixed effects. All standard errors are clustered by sublocation.

Table C5: Quality markers: control vs pure control

	(1)	(2)	(3)
	Has all quality markers	purity	germ_rate
Baseline Surveys Administered	-0.014	0.057	3.242*
	(0.058)	(0.039)	(1.773)
Observations	258	169	169
County FE	Yes	Yes	Yes
Control Mean	0.616	99.86	92.01

This table examines the influence of baseline survey activities on quality of seeds offered to secret shoppers that posed as uninformed buyers. The sample includes control and pure control markets in the main season in 2021. 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). Specifications control for county fixed effects. All standard errors are clustered by sublocation.

Table C6: Prices: control vs pure control

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

This table examines the influence of baseline survey activities on price of seeds offered to secret shoppers that posed as uninformed buyers. The sample includes control and pure control markets in the main season in 2021. All standard errors are clustered by sublocation. The dependent variable is the price paid (in Kenyan shillings) for one 2-kilogram packet of hybrid maize seeds.

D Appendix D: Extra Tables and Figures

Table D1: Quality markers versus price

	Price paid for 2kg hybrid	Price paid for 2kg hybrid	Price paid for 2kg hybrid
Has all quality markers	1.1	-0.5	2.5
	(2.3)	(3.1)	(3.3)
Observations	847	372	475
Sample	All Markets	Treated Markets	Control Markets

The sample includes markets in the main season in 2020, the short season in 2020, and the main season in 2021. Prices are measured in Kenyan shillings. All regressions include county, planting season, and brand fixed effects.

Table D2: Treatment delivery

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

This table shows the numbers of flyers distributed and residents trained directly as part of the information campaign that was implemented at treatment sites starting in February 2020.

Table D3: Attrition

	(1)
	Attritted
Treated	-0.01
	(0.02)
Constant	0.15***
	(0.01)
Observations	2193

This table tests for differential attrition across treatment and control households. The sample consists of all enrolled households, and the dependent variable equals one if kilograms of maize harvested for the main planting season was not successfully recorded in the follow-up survey activities. The dependent variable equals zero if kilograms of maize in the main planting season was captured in the follow-up surveys.

Table D4: Baseline balance

Variable	Control Mean	Treatment Mean	Difference
Market Area Population	138.7	138.5	-0.262
			(9.906)
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)	908.2	865.3	-42.97
			(35.346)
Bought Hybrid Maize Seeds (2019 main season)	0.965	0.964	-0.00100
			(0.017)
Planted maize (2019 short season)	0.267	0.254	-0.0130
			(0.029)
Joint F test – p-value = .515			

The sample includes 2431 households surveyed at baseline. The number of seed sellers was measured at baseline according to self-reports by sellers as to whether they are seed sellers. Home quality index captures the quality of roof, floor, and wall materials for a household's main residential building. P-value for joint test for differences between treatment groups is 0.51. For all tests for differences between the treatment group and the control group, we control for county fixed effects; counties are the level at which site selection was stratified. Standard errors are clustered at the sub-location level, which was the unit of randomization.

Table D5: 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 sample includes 467 seed packets that were purchased by secret shoppers from control sites and tested in a lab by KEPHIS staff. This includes packets purchased during the main season in 2020, the short season in 2020, and the main season in 2021. The dependent variable in all columns is the estimated percentage of material that are whole maize seeds (in percentage points). Each of the six independent variables equals 1 if the indicated quality marker is present and 0 otherwise.

Table D6: 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 sample includes 467 seed packets that were purchased by secret shoppers from control sites and tested in a lab by KEPHIS staff. This includes packets purchased during the main season in 2020, the short season in 2020, and the main season in 2021. The dependent variable in all columns is the percentage of seeds that germinated in the lab (in percentage points). Each of the six independent variables equals 1 if the indicated quality marker is present and 0 otherwise.

Table D7: Remote markets have greater potential gains in seed quality

	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 quality markers" equals 1 if all quality markers are present and 0 otherwise

Table D8: Households knowledge (by education)

	(1) Able to verify	(2) Expiration mentioned	(3) E-verification mentioned
Treat x No Primary Educ	0.080***	-0.015	0.118***
	(0.030)	(0.027)	(0.031)
Treat x At Least Primary Educ	0.277***	0.096***	0.337***
	(0.032)	(0.026)	(0.031)
Observations	2018	2002	2002
Control Mean	0.18	0.15	0.29

The table shows the effects of information treatment on self-reported usage of markers and respondent recall of knowledge included in the information treatment. The sample includes respondents that planted hybrid maize in the main growing season of 2020. The dependent variables are (1) whether the respondent indicated they were able to verify seed quality for the main planting season in 2020 (=1 if yes), (2) whether the respondent discussed expiration date as a way to verify seed quality (=1 if yes), and (3) whether the respondent discussed e-verification as a way to verify seed quality (=1 if yes). Measures of knowledge and ability to verify quality were collected via phone survey in July to August 2020, which was approximately 5 to 6 months after the information treatment was implemented. 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. Standard errors are clustered by sublocation.

Table D9: Prices for hybrid maize seed, paid by households

	(1) Avg price per kg	(2) Avg price per kg from local market	(3) Avg price per kg from elsewhere
Treated	2.214	-1.611	4.448
	(2.448)	(2.925)	(3.338)
Observations	2396	1243	1123
Control Mean	219.23	223.79	219.26
Treatment Effect (%)	1.01	-0.72	2.03

This table shows treatment effects on hybrid maize seed prices paid by households. The sample includes households in the main season of 2020 and the short season of 2020. The dependent variable is the average price paid per kilogram for hybrid maize seeds (in Kenyan shillings). Specifications control for gender, age, and education of household head, and county and planting season fixed effects. Standard errors are clustered by sublocation.

Table D10: Household self-reported germination rate

	(1)
	Germination rate
Treated	-0.79
	(0.72)
Observations	2399
Control Mean	85.04
Treatment Effect (%)	-0.93

This table shows treatment effects on households' self-reported germination rates. Results include households in both the main season in 2020 and the short season in 2020. The dependent variable is the self-reported germination rate expressed in percentage points (weighted average across all maize varieties planted by the household in the given planting season). Specifications control for gender, age, and education of household head, dependent variable measured in the 2019 main season, and county and planting season fixed effects. Standard errors are clustered by sublocation.

Table D11: Treatment effects on agricultural production, excluding households that purchased seeds prior to treatment

	(1)	(2)
	Kgs harvested	Yield
Treated	65.92**	52.91*
	(30.13)	(29.38)
Observations	3470	2112
Control Mean	508.10	840.21
Treatment Effect (%)	12.97	6.30

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, excluding any households in the main season who reported purchasing some seeds prior to the baseline survey. Yield is measured as kilograms harvested per acre. Specifications control for gender, age, and education of household head, dependent variable measured in the 2019 main season, and county and planting season fixed effects. All standard errors are clustered by sublocation.

Table D12: Treatment effects on agricultural production (post-double selection)

	(1) Yield
Treated	61.33**
	(28.28)
Observations	2355
Control Mean	853.33
Treatment Effect (%)	7.19

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. Estimates are produced using a post-double selection LASSO approach to select control variables, as described in Belloni et al. (2013). The procedure selects controls among those discussed in subsection 5.2, baseline characteristics shown in Table D4, and two and three way interactions.

Table D13: Treatment effects on agricultural production, 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
Control Mean	854.24	847.00
Treatment Effect (%)	14.53	7.21

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, excluding any households that indicated in the baseline survey that they purchased seeds from a close personal relation in the previous main season. Yield is measured as kilograms harvested per acre. Specifications control for gender, age, and education of household head, dependent variable measured in the 2019 main season, and county and planting season fixed effects. All standard errors are clustered by sublocation.

Table D14: Treatment effects agricultural production, excluding households that previously purchased on credit

	(1) Kgs harvested	(2) Yield
Treated	41.17	49.28
	(38.70)	(35.47)
Observations	2581	1708
Control Mean	553.34	820.14
Treatment Effect (%)	7.44	6.01

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, excluding any households that indicated in the baseline survey that they purchased seeds on credit in the previous main season. Yield is measured as kilograms harvested per acre. Specifications control for gender, age, and education of household head, dependent variable measured in the 2019 main season, and county and planting season fixed effects. All standard errors are clustered by sublocation.

Table D15: Household complementary inputs

	(1) Used Fertilizer
Treated	-0.00
	(0.01)
Observations	2871
Control Mean	0.93
Treatment Effect (%)	-0.24

This table shows treatment effects on household use of fertilizer. The dependent variable equals 1 if the household indicated that they used fertilizer (at all) for maize during that planting season. The sample includes households in both the main season in 2020 and the short season in 2020. Specifications control for gender, age, and education of household head, dependent variable measured in the 2019 main season, and county and planting season fixed effects. Standard errors are clustered by sublocation.

Table D16: Seller entry/exit (surveyed sellers only)

	(1) Sold seeds in main season	(2) Sold seeds in short season
Treated	-0.002	-0.065
	(0.030)	(0.050)
Observations	425	425
Control Mean	0.93	0.21
Treatment Effect (%)	-0.19	-30.62

This table shows treatment effects on seller entry and exit in the local market. The sample includes sellers recruited for the baseline survey. In column 1, the dependent variable equals one if the seller sold seed during the main season of 2020. In column 2, the dependent variable equals one if the seller sold seed during the short season of 2020. Specifications control for county fixed effects. All standard errors are clustered by sublocation.

Table D17: Seller entry/exit (surveyed sellers only; by baseline profit)

	(1) Sold seeds in main season	(2) Sold seeds in short season
Treat x Below Median Profit	-0.040	-0.084
	(0.043)	(0.055)
Treat x Above Median Profit	0.020	-0.052
	(0.028)	(0.058)
Observations	425	425
Control Mean	0.93	0.21

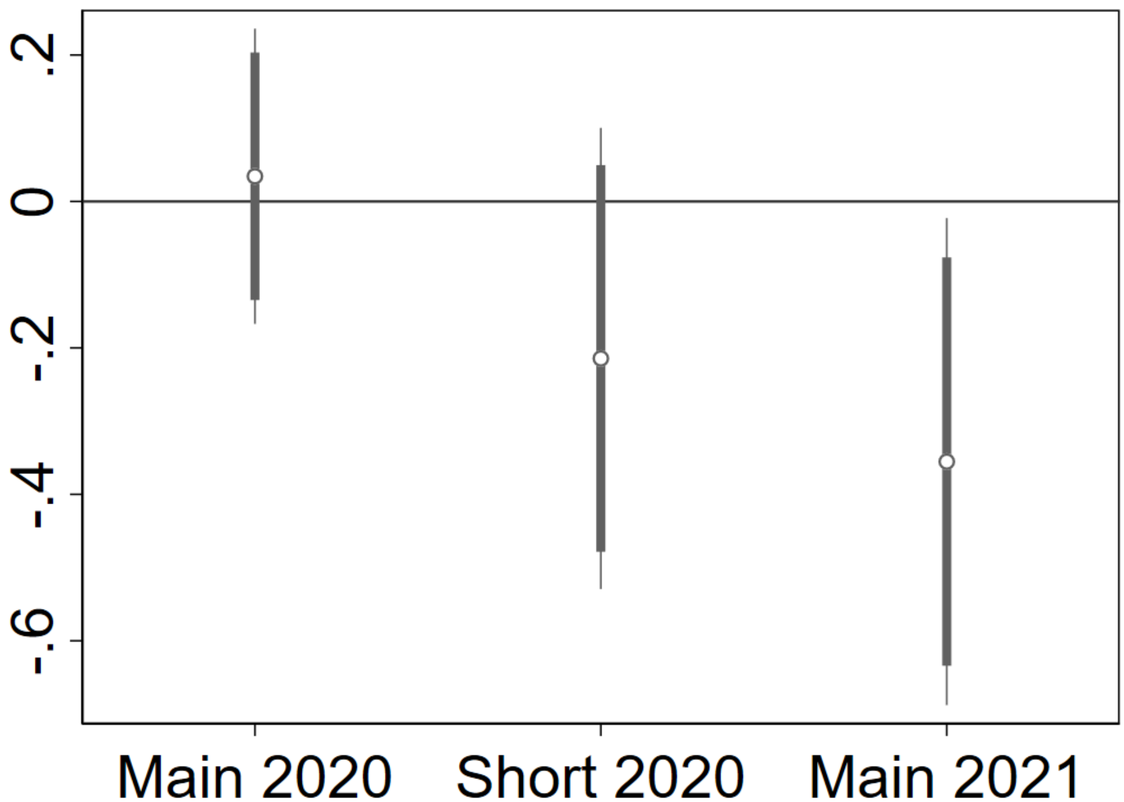
This table shows treatment effects on seller entry and exit in the local market. The sample includes sellers recruited for the baseline survey. In column 1, the dependent variable equals one if the seller sold seed during the main season of 2020. In column 2, the dependent variable equals one if the seller sold seed during the short season of 2020. In the estimation equation, a binary treatment indicator is interacted with either a dummy indicating that the business had below-median profits in the previous main season, or a dummy indicating it had above-median profits. Specifications control for county fixed effects. All standard errors are clustered by sublocation.

Table D18: Prices for hybrid maize seed

	(1) Price paid for 2kg hybrid	(2) SD of price	(3) Range of price
Treated	0.414	5.352	8.162
	(6.466)	(4.501)	(6.823)
Observations	354	229	229
Control Mean	480.71	26.94	41.20
Treatment Effect (%)	0.09	19.86	19.81

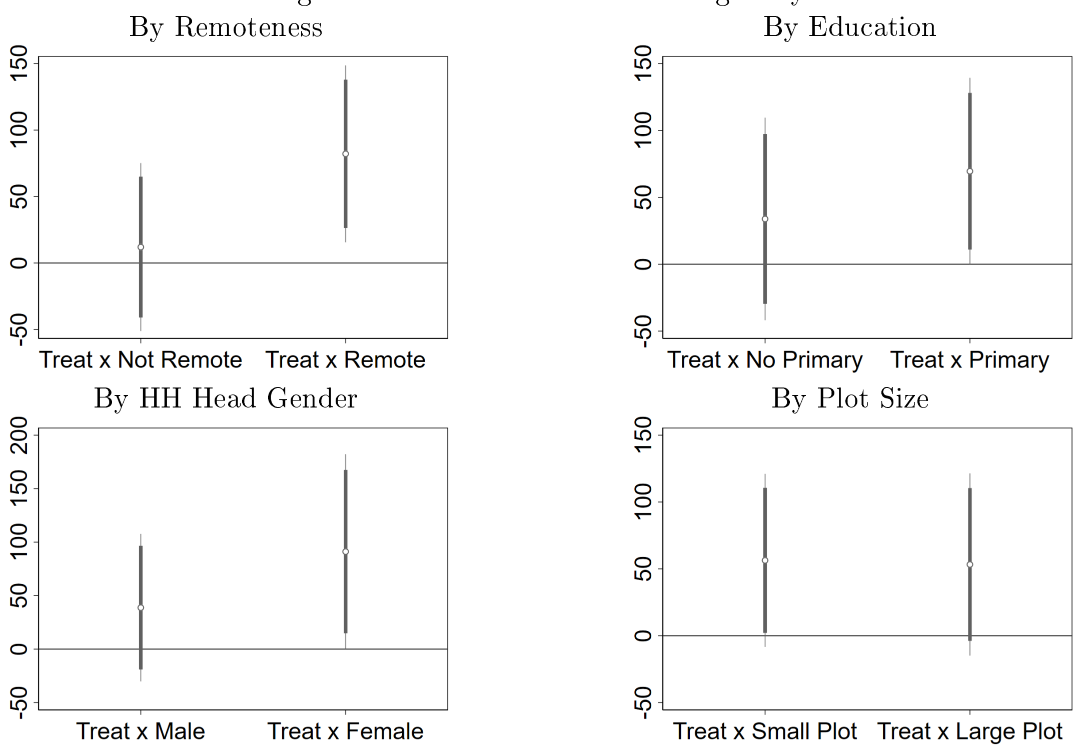
This table shows treatment effects on levels and dispersion of hybrid maize seed prices as observed by secret shoppers. The sample includes sublocations from the short season of 2020, and the main season of 2021. In column 1, the dependent variable is price paid in Kenyan shillings for a 2kg packet of seeds. In columns 2 and 3, the dependent variables are the standard deviation and range of the price paid. All standard errors are clustered by sublocation.

Figure D1: Treatment effects on seller entry/exit, by planting season



This figure illustrates treatment effects on number of seed sellers in each planting season. The dependent variable equals the number of sellers present at the market and offering hybrid seeds for sale, as observed by secret shoppers. A binary indicator for treatment is interacted with dummies for each planting season. The specification controls for county and planting season fixed effects. The thick and thin portions of each line represents 90% and 95% confidence intervals. Standard errors are clustered by sublocation.

Figure D2: 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. Each line represents 90% and 95% confidence intervals.

References

Abate, Gashaw T., Tanguy Bernard, Alain de Janvry, Elisabeth Sadoulet, and Carly Trachtman, "Introducing quality certification in staple food markets in Sub-Saharan Africa: Four conditions for successful implementation," *Food Policy*, 2021, *105* (C).

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.

Barahona, Nano, Cristóbal Otero, and Sebastián Otero, "Equilibrium Effects of Food Labeling Policies," *Econometrica*, 2023, *91* (3), 839–868.

Barriga, Alicia and Nathan Fiala, "The supply chain for seed in Uganda: Where does it go wrong?," *World Development*, 2020, *130* (C).

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "Inference on Treatment Effects after Selection among High-Dimensional Controls†," *The Review of Economic Studies*, 11 2013, *81* (2), 608–650.

Bennett, Daniel and Wesley Yin, "The Market for High-Quality Medicine: Retail Chain Entry and Drug Quality in India," *The Review of Economics and Statistics*, March 2019, *101* (1), 76–90.

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.

Bulte, Erwin, Salvatore Di Falco, Menale Kassie, and Xavier Vollenweider, "Low-Quality Seeds, Labor Supply and Economic Returns: Experimental Evidence from Tanzania," *The Review of Economics and Statistics*, 02 2023, pp. 1–33.

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.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, "The Value of Regulatory Discretion: Estimates From Environmental Inspections in India," *Econometrica*, 2018, *86* (6), 2123–2160.

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.

Hotz, V. Joseph and Mo Xiao, "The Impact of Regulations on the Supply and Quality of Care in Child Care Markets," *American Economic Review*, August 2011, *101* (5), 1775–1805.

Hui, Xiang, Ginger Zhe Jin, and Meng Liu, "Designing Quality Certificates: Insights from eBay," NBER Working Papers 29674, National Bureau of Economic Research, Inc January 2022.

J., Stevens-P. Harris and J. Morris, "Keeping It Real: Combating the Spread of Fake Drugs in Poor Countries," 2009.

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

Jensen, Robert and Nolan H. Miller, "Market Integration, Demand, and the Growth of Firms: Evidence from a Natural Experiment in India," *American Economic Review*, December 2018, *108* (12), 3583–3625.

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

Kala, Namrata, Clare Balboni, and Shweta Bhogale, "Climate Adaptation," *VoxDevLit*, 2023, *7*, 3.

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.

Leland, Hayne E., "Quacks, Lemons, and Licensing: A Theory of Minimum Quality Standards," *Journal of Political Economy*, 1979, *87* (6), 1328–1346.

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.

Michuda, Aleksandr, Cristina Chiarella, and Oscar Barriga-Cabanillas, "Genetic Dilution erodes productivity: Exploring farmers' low adoption levels of improved maize in Ethiopia," Working Paper 2022.

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.

Ngila, Faustine, "Fight against fake agro-inputs spreads in Comesa bloc," *The Daily Nation*, May 2019.

Obura, Fredrick, "Kenyan Minister unveils anti-fake seed," *The Standard*, 2013.

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.

Okwakol, Sam, "Alarm as fake seed merchants come up with new ways of selling products," *AllAfrica*, November 2015.

Romaniuk, Jenni, "Comparing prompted and unprompted methods for measuring consumer brand associations," *Journal of Targeting, Measurement and Analysis for Marketing*, October 2006, *15* (1), 3–11.

Schneider, Marius, Vanessa Ferguson, and Nora Ho Tu Nam, "The real problem of fakes in Africa: enabling factors and advice to rights holders," *Journal of Intellectual Property Law & Practice*, 04 2020, *15* (5), 365–377.

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.

Waller, Jo, Kirsten McCaffery, and Jane Wardle, "Measuring cancer knowledge: Comparing prompted and unprompted recall," *British Journal of Psychology*, May 2004, *95* (2), 219–234.

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

World Bank, IFC, and MIGA, *World Bank Group Climate Change Action Plan 2016-2020* number 24451. In 'World Bank Publications - Books.', The World Bank Group, December 2016.