Filling a Niche? The Maize Productivity Impacts of Adaptive Breeding by a Local Seed Company in Kenya

Samuel S. Bird[a,*], Michael R. Carter[b,c,d], Travis J. Lybbert[b], Mary Mathenge[e], Timothy Njagi[f], Emilia Tjernström[g]

[a] *Luther College, Decorah, IA, 52101, USA.*
[b] *University of California Davis, Davis, CA, 95616, USA.*
[c] *National Bureau of Economic Research, Cambridge, MA, 02138, USA.*
[d] *University of Cape Town, Cape Town, 700, South Africa.*
[e] *Egerton University, Njoro, Nakuru, Kenya.*
[f] *Tegemeo Institute of Agricultural Policy and Development, Nairobi, Kenya.*
[g] *Monash University, VIC 3800, Australia.*

Abstract

This paper studies whether the absence of locally adapted seed varieties constrains the productivity and incomes of farm households residing in small, agro-ecological niches. We empirically examine the disruption of the maize seed market in Western Kenya that took place when public sector foundation breeding and social impact investment capital came together and allowed a local seed company to expand and target a niche area with adaptively-bred maize varieties. The three-year RCT reveals that these seed varieties increased farmer yields and revenues, both for better-resourced farmers (who used non-adapted hybrids and fertilizer prior to the intervention) as well less well-resourced farmers (who did not). This theoretical and empirical evidence suggests news ways for thinking about seed systems in areas typified by high levels of agro-ecological heterogeneity.

Keywords: technology adoption, innovation, randomized controlled trial, Kenya, maize, seed systems, (Classification codes: O12, O13, Q12, Q16)

---

[*] Corresponding author
Email address: `birdsa01@luther.edu` (Samuel S. Bird)

Preprint submitted to Elsevier

1. Introduction

Since 1960, agricultural productivity in sub-Saharan Africa has grown much more slowly than in other regions in the developing world. In 1960, cereal yields in sub-Saharan Africa were just below those of yields in Asia and Latin America. By the early 1990s, this yield gap had more than tripled to 0.6 tons per acre, and by 2017, the gap had doubled again to over 1.2 tons per acre (Carter et al., 2021). The gap is largely attributable to the region's failure to adopt improved green revolution cereal varieties and other complementary inputs (Evenson and Gollin, 2003); fewer than half of Sub-Saharan African farmers employ improved varieties, in contrast to near universal adoption elsewhere. The puzzle of this persistently low adoption rate has motivated a large literature, which identifies a range of constraints to adoption, ranging from behavioral biases and other internal or psychological constraints (Abay et al., 2017; Duflo et al., 2011), to information (Carter et al., 2021), risk (Dercon and Christiaensen, 2011; Karlan et al., 2014), and biophysical and other external resource constraints (Marenya and Barrett, 2009; Suri, 2011). This paper explores an alternative, and ultimately complementary, explanation for this puzzle, highlighting supply-side constraints to the adoption of green revolution varieties appropriate for adoption in sub-Saharan Africa.

In many ways, Kenya is an exception to the sub-Saharan African pattern of low adoption of improved cereal varieties. By the early-1990s, if not earlier, a large majority of Kenyan farmers had in fact adopted hybrid maize varieties. However, this high average rate of adoption obscures important heterogeneity across regions of Kenya. As shown in Figure 1, in Kenya's two largest maize growing regions, the highland and transitional zones, adoption rates have been above 75% since the 1970s. These hybrid adoption rates contrast with the mid-altitude zone, where this rate has barely crept above 25% in the years since 1970.[1] This mid-altitude zone is relatively small, constituting only 11% of Kenya's maize area, and yet is home to substantial numbers of Kenya's poor farm households (Hassan, 1998).

Possible explanations for this persistent pattern of low hybrid adoption in Kenya's mid-altitude zone include that it is not biologically possible to develop hybrids that perform well

---

[1] Throughout this paper we use the agro-ecological zones for maize production in Kenya developed by the International Center for Improvement of Maize and Wheat (CIMMYT) (Hassan, 1998). These zones are defined by length of the maize growing season, which itself depends in large part on temperature (degree days), as well as rainfall and altitude. This paper focuses on the main zones for maize production: the highland tropical (henceforth, highland) zone composed of farms at or above 1600-2900 meters in altitude, the moist transitional (henceforth, transitional) zone from 1200-2000 meters in altitude, and the moist mid-altitude (henceforth, mid-altitude) zone from 1110-1500 meters in altitude. Figure 1 is constructed based on the best available data from these zones, which came from Gerhart (1975), Hassan (1998), TAMPA2 (2004), TAPRA (2010), and the baseline survey for our study.

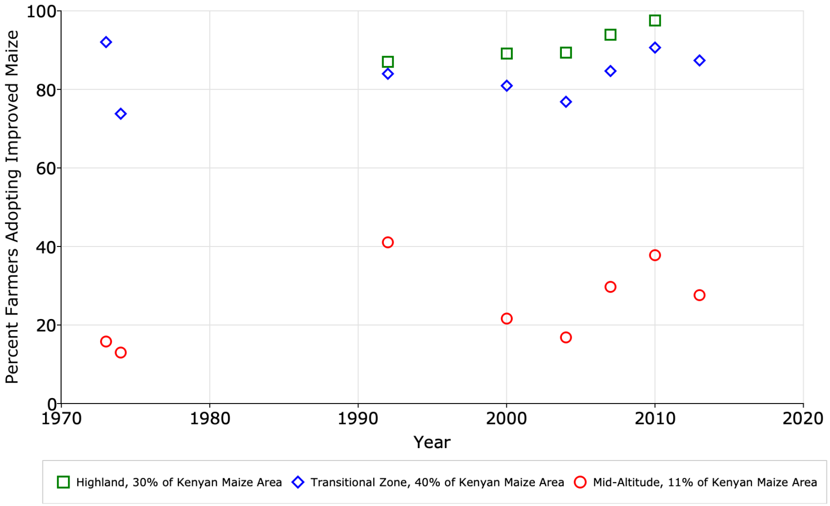


Figure 1: Persistently low adoption of improved maize in Kenya's mid-altitude zone.

and are profitable to adopt in the area given its soils, altitude and growing season. If correct, this explanation would imply that an effort to develop and introduce hybrids adaptively bred for this area would not result in any demonstrable yield or economic gains for farmers. A second explanation is that local seed adaptation is possible, but that a confluence of supply and demand factors sustain a market equilibrium in which no investment in seed adaptation takes place. If this perspective is correct, then disruption of the constraints that sustain this equilibrium would be expected to lead to substantial gains for producers.

Empirically distinguishing between these explanations, and better understanding the limits facing farmers in Kenya's mid-altitude and perhaps other small agro-ecological zones, is unfortunately not straightforward. While we are unable to experimentally manipulate the entrance of new seed companies across a sample of niche geographies, we are able to experimentally test the key implication of these competing explanations by taking advantage of the rapid expansion of a local seed company focused on creating maize varieties specifically adapted for western Kenya, including the mid-altitude zone.

Prior to the time of our experiment, the seed company in question, Western Seed Company, benefited from a public-private partnership with the International Center for Improvement of Maize and Wheat (CIMMYT) to freely use CIMMYT's parent seed lines to innovate varieties specifically adapted for the mid-altitude zone. In the mid-altitude zone, the key attribute needed for locally adapted varieties is the ability to mature during the short growing seasons. However, Western Seed's seed breeding and replication operations were limited by capital constraints, similar to other local seed companies in the region (Langyintuo et al., 2010). Western Seed's capital constraint was relaxed when a social impact investor (Acumen) provided Western Seed the funds to radically expand its breeding and seed multiplication operations. This expansion in turn permitted the research team to conduct a randomized control trial to measure the farmer impacts of introducing Western Seed's locally adapted varieties.

In line with the second explanation above that, absent disruption, demand and supply factors sustain a low innovation equilibrium, the RCT reveals that the introduction of Western Seed varieties caused substantial yield gains for maize farmers in the mid-altitude zone. Outside of this zone, we find that the new seed varieties provided by Western Seed performed no differently than the improved varieties already available in the market. Digging deeper, we note that Figure 1 suggests the presence of substantial heterogeneity in the mid-altitude zone, with a minority of farmers using hybrid maize, while the majority do not. Guided by a formal model of variety adoption, we unpack these average treatment effects and explore the differential impact of the introduction of Western Seed varieties on these two groups. For farmers who historically did not use improved maize varieties, our intention to treat impact estimates reveal that the availability of Western Seed varieties increases yields by 21% on average, a large effect despite these farmers using little to no complementary inputs like fertilizers.[2] However we would expect farmers who historically used improved seeds to have done so in part because they had the resources to invest in complementary inputs like fertilizers. We find this to be true empirically, and that these farmers realize an even larger average yield gain of 47% due to the availability of Western Seed varieties.

Stepping back from Kenya, sub-Saharan Africa is known to be comprised of a wide-variety of different agro-ecologies. We cannot in this paper pin down the extent to which the broader sub-Saharan African pattern of low use of green revolution technologies can be attributed to small or niche agro-ecologies for which locally adapted improved varieties do not exist. But our results point to a public-private seed sector model that potentially offers benefits across the wealth spectrum of African cereal farmers.

The remainder of this paper is structured as follows. Section 2 presents a model of seed variety choice by low farm households both before and after the introduction of locally

---

[2] With net compliance rates 14% for this group, impacts on those who actually adapted the seeds are substantially higher, as discussed later in the paper.

adapted varieties and provides insights for the design of the subsequent empirical analysis. Section 3 introduces the western Kenya study area and lays out the design for the RCT made possible by the capacity expansion of Western Seed. Section 4 presents average treatment effects for the mid-altitude zone as well as for other zones included in our study. Section 5 undertakes the key heterogeneity analysis, identifying the impacts within the mid-altitude zones on farmers that had and had not been prior users of (non-locally adapted) hybrid seeds. The final section concludes with reflections on implications for seed systems in areas with substantial agro-ecological heterogeneity.

2. Seed Variety Choice in a Niche Agro-Ecological Zone

As a prelude to our analysis of the experimental introduction of locally adapted maize seeds into Kenya's mid-altitude zone, this section uses a model of a small farm household to provide microeconomic intuitions along two dimensions that help situate and structure our analysis. The first concerns the demand-side constraints that might create the stable baseline scenario scene in Figure 1 in which only a modest minority of mid-altitude farmers adopted the then available hybrid maize seeds that had been bred for higher altitude environments. The second, and related dimension concerns the the heterogenous impacts that we would expect to see upon the introduction of the locally adapted seed technology into this baseline scenario.

The farm household model rests on the assumption that maize is produced according to the following function:

$$y^v(f^v) = \theta\left(\alpha_0^v + \alpha_1^v f^v\right) \quad \forall f < f^o,$$

where the variety indicator superscript v takes on the value of r for retained local seed, n for non-locally adapted hybrid (NLA) variety, and a for locally adapted hybrid variety. The term f^v is the per-acre intensity of fertilizer applied to variety v and f^o is the agronomically optimal fertilizer rate.[3] θ is a multiplicative shock term, with support $\left[0 - \overline{\theta}\right]$ and $E\left(\theta\right) = 1$. Assigning a price of p_m for maize, a farmer who devotes H^v hectares of land to variety v will earn the following net income (i.e., less input costs):

$$Y^v = H^v\left(\left(\theta\alpha_0^v p_m - s^v\right) + \left(\theta\alpha_1^v p_m - p_f\right) f^v\right)$$

---

[3] In reality, returns to fertilizer, α_1^v, vary with soil quality (Tjernstrom 2017) and farmer skill (Laajaj and Macours, forthcoming). Adding in these additional dimensions of farm or farmer heterogeneity would add complexity, but little additional insight.

where s^v is the per-acre cost of seed for variety type v, and p_f is the price of fertilizer. Appendix A discusses the logic of this linear specification, which is based on the assumptions that maize exhibits increasing returns to nutrition at low nutrient levels and that farmers optimally manage fertilizer application rates.

Using this notation, we assume the following stylized characterizations of the three seed technologies.

- *Retained local variety* (r)

  The per-hectare cost of local seeds, s^r, is low (since farmers can save grain from the previous harvest). Fertilizer application is not profitable in expectation $(p_f > \alpha_1^r p_m)$.

- *Non-Locally adapted hybrid variety* (n)

  The cost of NLA seeds is higher than local seeds $(s^n > s^r)$, and while this variety is fertilizer-responsive $(\alpha_1^n p_m > p_f)$, its unsuitability to the local environment means that it is less profitable than local varieties without complementary fertilizer application, i.e. $((\alpha_0^n p_m - s^n) < (\alpha_0^r p_m - s^r))$.

- *Locally adapted hybrid variety* (a)

  Also more costly to the farmer than retained seeds, we assume that these varieties are no more costly than the NLA hybrid $(s^n \geq s^a > s^r)$.

Building on this stylized production structure, the remainder of this section models the adoption of improved seed varieties both before and after the introduction of locally adapted hybrids.

2.1. Seed Variety Choice in the Baseline Scenario

We first consider the optimal choice of seed variety when farmers can only choose between varieties r and n. With reference to the RCT described in greater detail below, this choice corresponds to the baseline scenario. Appendix B writes down a formal, two-period optimization problem for a risk averse farmer with an initial endowment of planting season cash-on-hand, z_0, that she must use to finance the purchase of seeds and fertilizer and cover family consumption expenses over the growing season. In addition, the farmer can save a portion of her initial endowment as a buffer against future income shocks, using an informal savings technology that has an interest rate less than the farmer's rate of discount. The farmer may exhibit a degree of skepticism about hybrid varieties and behave as if the production parameters for the hybrid technologies are a fraction λ of their true value, $0 < \lambda \leq 1$.

Such technological skepticism could reflect prevalent suspicions about the quality of agricultural inputs (Bold et al., 2017) or a more general lack of experience with hybrids and of opportunities to learn their true value on-farm or through one's social network.[4] Finally, the model assumes that there are minimum scales of adoption based on the assumptions that farmers will not buy a partial (opened) bag of seed or fertilizer because of concerns about input counterfeiting.[5]

As a way to quickly gain intuition about this model, Figure 2 displays the results from numerical analysis of the optimization model for farmers spread across the two-dimensional space defined by cash-on-hand (the x-axis) and technology skepticism (the y-axis). All farmers are assumed to have an identical 2 acres of farmland on which they can cultivate maize. Table A1 in the appendix lists the specific price, productivity, and preference parameters which have been chosen to be match the reality in Kenya's mid-altitude zone. The NLA hybrids cost 10 times as much as retained seeds and are not profitable to adopt without fertilizer given the Table A1 parameter values. A kilogram of nitrogen fertilizer applied to the NLA hybrid returns 20 kilograms of additional maize. This figure is at the low end of the estimates reported in the review paper by Jayne et al. (2018), and is intended to capture what it means for a variety to be non-locally adapted. In financial terms, a dollar of fertilizer returns \$1.20 worth of maize under the NLA technology under our data-based price assumptions. Note that this rate of return is 4-times the farmer's assumed 5% rate of discount. Adopting the NLA technology at the minimum scale (2 kg of seed and 10 kg of fertilizer) would cost \$88. Expected gross agricultural income for the two-acre maize farmer using exclusively retained seeds would be \$240, with only \$8 in total planting (seed) costs.

In Figure 2, the solid (black) contour line divides the endowment-skepticism space into two groups: *Group 1* in this baseline scenario does not adopt NLA hybrids, whereas *Group 2* adopts them and applies fertilizer. As can be seen, only farmers with more than \$325 of planting season cash-on-hand will adopt the NLA technology, even if they are not skepti-

---

[4] As modeled here, skepticism could emerge from a lack of learning as captured by models of learning about productivity (Besley and Case, 1993 and Munshi, 2004), in contast to target input models of learning (Foster and Rosenzweig, 1995 and Conley and Udry, 2010). For example, Munshi (2004) underscores how local heterogeneity slowed the diffusion of improved varieties in India because it hampered social learning about the value of these new technologies. Empirically, there is ample evidence of such skepticism in settings similar to ours. In their study of an input subsidy program designed to encourage experimentation with improved varieties and fertilizer, Carter et al. (2021) find that in their baseline data farmers massively under-estimate the returns they would receive from using improved inputs compared to what they actually experienced once they were induced to try out the new inputs.

[5] In our study area, Western Seed and other companies faced such significant problems of counterfeit seeds (unscrupulous individuals would collect used seed bags and refill them with local seeds) that they began enclosing tickets that could be used to certify seeds authenticity through an SMS-based message system.

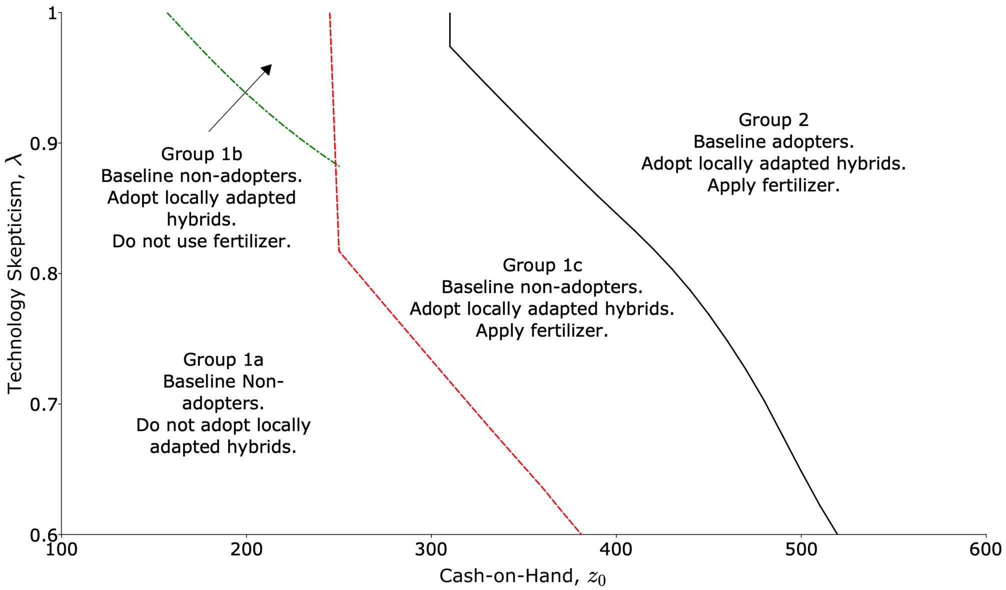


Figure 2: Hybrid seed adoption patterns (Case 1 parameter values).

cal ($\lambda = 1$). As skepticism increases (*i.e.*, as λ decreases), farmers do not adopt the NLA technology unless cash holdings are even higher. A farmer's reluctance to invest is of course being driven both by borrowing constraints (investment comes at the cost of consumption whose marginal utility is high for low wealth farmers) and the fact that they must autarkically manage risk through informal savings. Optimal savings rates vary from 5-20% of the cash endowment, despite the fact that the farmer's discount rate exceeds the assumed non-positive rate of return on informal savings. For those who do optimally adopt, their agricultural production on average increases by a factor of two, although the financial gains are more modest.

While this fairly simple model could be extended to consider dynamic issues (especially the dynamics of learning about new technologies), it does provide insight into the forces that likely keep households both poor and skeptical when facing the modest returns offered by an NLA technology. Depending on the distribution of the population across the $z_0 - \lambda$ space, the number of people adopting the NLA hybrids could be relatively modest, as has been the case for the last 50 years in Kenya's mid-altitude zone. Note that even a temporary input subsidy program of the sort studied by Carter et al. (2021) might have modest impact in the presence of NLA technology even if it successfully induced the kind of learning and reduction of skepticism found by those authors. In summary, the effective size of a niche market can be made even smaller by the confluence of poverty and skeptical inexperience

Table 1: Benefits from introducing an adapted hybrid in a niche agro-ecological zone.

	Adapted Hybrid More Fertilizer Responsive ($\alpha_1^a > \alpha_1^n$)	**Non-Adapted Hybrid More Fertilizer Responsive** ($\alpha_1^a < \alpha_1^n$)
Adapted Hybrid Outperforms Retained Variety without Fertilizer ($\alpha_0^a p_m - s^a > \alpha_0^r p_m - s^r$)	*Case 1* (Groups 1 & 2 Benefit)	*Case 2* (Group 1 Primarily Benefits)
Retained Variety Outperforms Adapted Hybrid without Fertilizer ($\alpha_0^a p_m - s^a < \alpha_0^r p_m - s^r$)	*Case 3* (Group 2 Primarily Benefits)	*Case 4* (Neither Group Benefits)

Group 1 farmers are those who do not adopt hybrids in the baseline scenario whereas Group 2 farmers do.

with improved technologies.

2.2. Technology and Variety Choice Following the Introduction of Locally Adapted Hybrids

So how might the introduction of locally adapted hybrids operate in this environment where the farming community is divided between those who did and did not adopt NLA hybrids bred for other agro-ecological zones? Table 1 considers four possible cases depending on the impact of the local adaptation on the key productivity parameters, α_0^a and α_1^a. In Case 1, the locally adapted variety outperforms the retained variety even when fertilizer is not used, and it also outperforms the NLA variety when fertilizer is used. In this case, introduction of the technology would benefit both Group 1 and Group 2 farmers. Case 4 is the opposite case, with the locally adapted variety outperforming neither retained nor NLA varieties under these conditions. The off-diagonal cases (2 and 3) are where the the locally adapted variety outperforms one type of seed-fertilizer combination, but not the other. The experimental introduction of locally adapted varieties to both Group 1 and Group 2 farmers will allow us to ultimately distinguish between these cases when new, locally adapted varieties were introduced into Kenya's mid-altitude zone.

As a prelude to that empirical analysis, we use the model in Appendix B to explore the expected impact of the introduction of a doubly successful (Case 1) locally adapted hybrid variety. Figure 2 displays the result of the optimization analysis for this Case 1 scenario in which the locally adapted hybrid outperforms local seeds without fertilizer and exhibits a higher marginal return to fertilizer than the NLA hybrid. The specific parameter values used for the numerical analysis assume that the gain without fertilizer over retained seeds is 87% in expectation for the locally adapted variety, and that the returns to fertilizer with this variety are 25% higher than returns to fertilizer using the NLA hybrid.[6]

Under this scenario, the portion of the space where farmers optimally rely on retained seeds shrinks to the area to the south and west of the broken line contour lines. This group, labeled Group 1a in the figure, is still too poor, skeptical and risk-constrained to adopt the locally adapted variety despite its relatively modest entry cost of \$40. To the north and east of Group 1a are two sets of farmers who adopt the locally adapted hybrid, with the more capital-constrained Group 1b adopting the improved seeds, but not purchasing any fertilizers. The introduction of the locally adapted hybrid induces Group 1c to both adopt the variety and apply fertilizer. Compared to their baseline state and assuming optimal adoption behavior, these groups would experience production increases of 10% (Group 1b) and 80-100% (Group 1c) under these parameter values.[7] Expected net revenue (value of production minus seed and fertilizer costs) increases by about 8% for Group 1b and about 65% to 70% for Group 1c. Note that these farmers experience large costs increases as they switch to much more expensive hybrid seeds and, in some cases, begin to purchase fertilizer.

Finally, under the Case 1 parameter values, Group 2 farmers would optimally adopt the locally adapted varieties, with expected production increases of 30-50%. Because these production impacts reflect optimizing behavior, underlying income and expected utility also increase for all groups. Expected net revenue increases for Group 2 farmers are slightly larger than the expected production increases as Group 2 farmers have no little change in costs when they switch from non-adapted to adapted hybrid (their costs only increase by the amount of the additional fertilizer they optimally purchase).

While these specific figures are of course artifacts of the numerical assumptions, they do illustrate the structural impact heterogeneity that could accompany the introduction of locally adapted varieties in an area like Kenya's mid-altitude zone. A simple estimate of the average treatment effect of the introduction of a locally adapted hybrid would be a data-weighted average of these separate effects, with the weights of course depending on the distribution of the farming population across the wealth-skepticism space. Note further that, under different parameter cases, average treatment effects will be pushed toward zero even when the technology is effective for a sub-set of farmers (*e.g.*, in Case 3 in Table 1,

---

[6] The parameters imply that fertilizer has average value-cost ratio of 1.6, which is in line with the Jayne et al. (2018) review of the literature on returns to fertilizer on maize in East Africa.

[7] Optimal adoption behavior is typified by only partial adoption, with improved seeds applied to a portion of the farm area and fertilizer intensity typically increasing with liquidity.

only Group 2 and a portion of Group 1c would benefit).

3. Empirical Context and Experimental Design

In the 1960s, the government of Kenya supported research and development of hybrid maize for the highland zone, which covers almost one-third of Kenya's maize-growing areas and has high agricultural potential (Gerhart, 1975; Hassan, 1998). Significantly less investment was made to develop varieties for other regions, including the mid-altitude zone. Seed market liberalization in the 1990s allowed the entry of major multi-national seed companies into the Kenyan market, although their focus was also on varieties tailored to the highland zone. Figure 3 reflects this regional pattern of seed adaptation and innovation. Data from the Kenyan Plant Health Inspectorate Service, which reports verified yields for certified varieties when grown at their recommended altitude and season length, shows that by 2010 there were a number of varieties yielding between 3.5 and 5 tons/acre in the highland zone.[8] In contrast, until the introduction of Western Seed varieties (indicated by red squares in the graph), there was only a modest number of registered varieties that yielded more than 3 tons/acre in the smaller, mid-altitude zone.

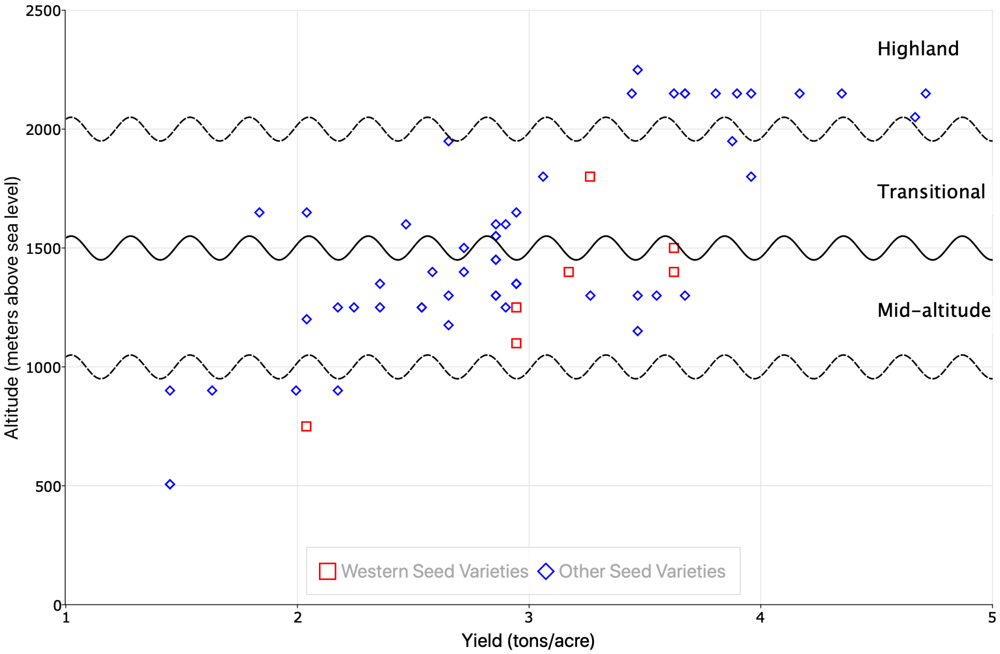
This regional pattern is consistent with a model in which the large national and multinational seed companies face fixed costs for adaptive breeding that discourage investment in varieties adapted for smaller, or niche, agro-ecological zones. While local seed companies based in these smaller zones would be expected to have lower fixed costs than outside firms (assuming they already know the particularities of their zone), they typically do not own the "parent" seed lines needed for successful adaptive breeding. They are also likely to face binding capital constraints.[9] Appendix C below further discusses these key features of varietal innovation which can explain persistence of a low innovation equilibrium in niche agro-ecologies.

The gap in hybrid maize adoption across regions of Kenya illustrated in Figure 1 was the focus of an in-depth study to chart a path for future maize research by CIMMYT (Hassan, 1998). The study recommended that CIMMYT develop parent lines for maize varieties that can mature during the shorter growing seasons typical of the mid-altitude zone of western

---

[8] As discussed in note 1 above, the different agro-ecological zones are not defined exclusively in terms of geography, and hence the lines demarcating zones in Figure 3 are displayed as fuzzy boundaries.

[9] Adaptive breeding is a multi-year process which requires firms to cover the costs of experimentation and variety identification as well as the waiting time for seed certification as agencies to grow and evaluate new varieties.



Source: Commercial recommendations for hybrid varieties from company websites among varieties registered by Kenya Plant health inspectorate Service (KEPHIS, 2018)

Figure 3: Yields and recommended planting altitude for registered maize varieties as of 2010.

Kenya. Subsequently, public investment in research and development, as well as reforms to seed markets, spurred private investment in innovation and product markets.[10] The shift in public-sector research by CIMMYT laid the groundwork for private firms, such as the Western Seed Company, to develop maize hybrids for the mid-altitude zone of western Kenya.

3.1. Expansion of Western Seed and the Randomized Controlled Trial

With its new varieties, the geographic footprint of Western Seed maize hybrids slowly expanded over time from the transitional zone into the mid-altitude zone near Lake Victoria.[11] However, Western Seed's seed multiplication and market expansion was constrained by its available capital (Partners). In 2008 and in 2010, two impact investment organizations (Pearl Capital Partners and Acumen Fund) made debt and equity investments in Western Seed totaling $3 million with the intention of rapidly tripling Western Seed's seed supply capacity. Within a few years, this new supply capacity was on-line, opening the door for

---

[10] In the early 1990s, the seed market in Kenya was liberalized.

[11] Tegemeo Institute's TAPRA data set allows us to see the expansion of Western Seed hybirds into the mid-altitude zone between the 2004 and 2010 TAPRA survey rounds (TAMPA2, 2004; TAPRA, 2010).

both Western Seed's geographic expansion and coincidentally creating a unique opportunity to establish an RCT around the introduction of locally adapted hybrids.

In partnership with Western Seed and Acumen, the research team established a research design that would allow identification of the impact of the introduction of Western Seed hybrids in new areas in western and central Kenya. Specifically, Western Seed had resources to establish 100 new demonstration plots at key points across these regions for the 2013 planting season. Each demonstration plot was designed to provide information to villages within a 5 to 10 mile radius of the plot and the sites were spaced with that distance in mind. At the research team's request, Western Seed identified 125 potential demonstration plot sites (25 more than they wanted) with the understanding that the team would randomly allocate up to 25 sites to a control group where no demonstration plots nor marketing would take place.

Figure 4 maps the study sites across central and western Kenya. The background shading on the mapping shows the approximate altitude ranges of highland areas (above 2000 meters), transitional areas (1500-2000 meters) and mid-altitude areas (below 1500 meters). While these altitude designations do not completely describe Kenya's agro-ecological zones (see footnote 1), they give insight into the geography of the study area. Each study site is marked by the symbol corresponding to its actual agro-ecological zone classification based on Hassan (1998).

Each site typically contains 3-5 villages within its zone of influence. The easterly, central sites all lie within the transitional zone, while the western sites are divided between the mid-altitude, transitional, and highland zones. For purposes of the analysis that follows, we group the two highland sites with the western transitional zone sites. The 36 sites randomly selected for the study were grouped into matched pairs based on physical proximity, altitude, and climate. One member of each pair was then allocated to the seed treatment and one to control status. The timeline in Figure 5 displays the full life of the intervention and study. A random sample of 50 farmers was selected for interview in each site, resulting in a total sample size of 1800 farm households.

The seed treatment consisted of three components. The first component was the establishment of demonstration plots for the 2013 main maize season so that nearby farmers could observe the performance of the Western Seed varieties.[12] The second component was

---

[12] The primary maize growing season in western Kenya stretches from March to September; some farmers

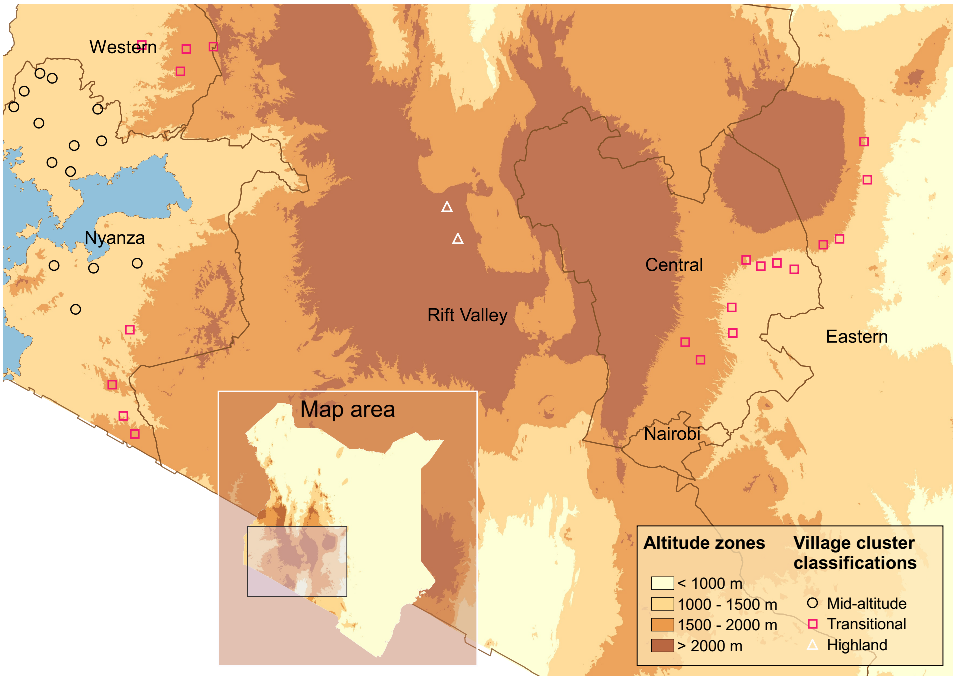


Figure 4: Study sample in western and central Kenya.

the provision of a small sample packet to farmers to try on their own farm for that same season. A Western Seed marketing representative visited each community to distribute the packets and provide further information on the Western Seed hybrids. Sample packets had 250 grams of seed, enough seed to plant one-fortieth of an acre. Farmers were asked to plant their trial packet separately and to keep track of its performance, which most did (see Tjernstrom (2017)). Given the small size of the seed packet, we expected it to inform farmers' future planting decisions, but not to influence their yields or income for the 2013 maize growing season. The third and final component was the offer to pre-order Western Seed hybrids and have them delivered to their village prior to the 2015 maize season. As discussed below, this third element was added following low uptake of Western Seed seed varieties in 2014.

In addition to the core seed treatment, we also implemented a fertilizer intervention for the 2014 maize season that gave fifty kilograms of high-quality fertilizer to randomly selected farmers in both treatment and control sites in the western study areas. At each site, the research team held a public lottery amongst survey participants, with half receiving

---

also plant a second maize crop in October, although this second season is typically less productive and receives fewer inputs from most farmers who plant it. The primary maize growing season in central Kenya is the October planting, with a less productive season from March to September.

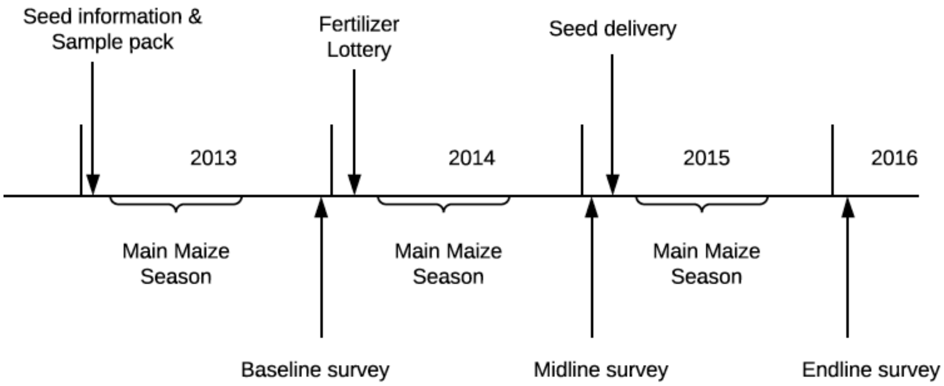


Figure 5: Study timeline.

fertilizer and the others a token gift of cell phone time. The motivation behind this ancillary intervention was to test the claim that the yield gains with Western Seed varieties are much greater for farmers applying fertilizer. We did not implement the fertilizer intervention in the central study area as baseline fertilizer use was quite high.

In summary, in the western study areas, assignment to the treatments randomly divided a total of 1200 farm households into four equally-sized groups:

1. A *Control* group;
2. A *Seed Only* treatment group that received Western Seed information and free 250 gram trial seed packets in 2013 and an option to have Western Seed varieties delivered (at cost) to their home in 2015;
3. A *Fertilizer Only* group that received fertilizer in 2014, but received no seed treatment;
4. A *Seed and Fertilizer* group that received each of the treatments received by groups 2 and 3.

In the central study area, 600 households were divided evenly between groups 1 and 2 only. As shown in Figure 5, baseline, midline, and endline surveys were held in both the central and western areas following the 2013, 2014, and 2015 main maize season harvests for western Kenya.

3.2. Baseline Characteristics and Compliance

Table 2 summarizes baseline characteristics of households in each zone: mid-altitude, transitional, and central. We restrict the sample to only those households that reported planting maize in each year of the study, as we do in the subsequent empirical analysis. We summarize baseline characteristics in levels for ease of interpretation.

The first column summarizes characteristics of households in the mid-altitude zone. Prior to Western Seed's expansion, on average households planted any hybrids in only 26% of previous 5 main maize seasons, and used fertilizer at the same low rate. Unsurprisingly, maize yields also are low, at 234 kilograms per acre. Yet maize is central to the livelihoods of smallholder households in the region; in our sample, households on average plant maize on 80% of their land. Maize is an important source of income and food for smallholder households. In the mid-altitude zone, average annual income per capita is low, at 22,800 Kenyan shillings, or approximately 584 PPP USD, with agriculture contributing 34% of total income, and maize contributing just under half of total agricultural income. This measure of low household income in the mid-altitude zone is substantiated by a separate asset-based index indicates that, on average, a household has almost a 1/3 probability of living on less than 1.25 PPP USD per person per day. Food insecurity is common among households in the sample, with almost 2/3 of households being food insecure at some point during the year.

The theory in Section 2 above suggests that, absent well-adapted hybrids, the small farming population will bifurcate into one group using exclusively local seeds and another group relying on hybrids and applying fertilizer. The second and third columns of Table 2 divide up mid-altitude respondents based on their pre-intervention use of hybrids. Hybrid users are those who had planted hybrids in at least 4 of the 5 pre-study main maize seasons. Hybrid non-users are those who planted hybrids in less than 4 of those seasons.[13] As can be seen in the table, the smaller hybrid-users group almost exclusively relies on hybrids, and applies fertilizer almost 60% of the time (and in almost all of the main maize seasons). The larger non-users group rarely uses hybrids and uses fertilizer at about one third rate of the users group. As expected, yields are much higher at baseline for the hybrid users group, at 342 kilograms per acre compared with the non-user group average of 211 kilograms per acre.[14] Yields for both of these group fall well short of the potential mid-altitude yields of 3-4 tons per acre shown in Figure 3. Incomes for the users group are about 50% higher than incomes for the non-users group, due in large part to greater non-agricultural income. This finding is consistent with the model in Section 2, which suggests that liquidity constraints

---

[13] Of the 700 households in the mid-altitude zone, we categorize 125 as users of hybrids in the seasons preceding the study. This includes four households that planted maize in fewer than 4 of the 5 pre-study main maize seasons, but planted hybrids in each of the seasons in which they planted maize.

[14] Two mid-altitude households have missing data for past hybrid use, which is why the observations for the non-user and user samples do not sum to the total number of observations in the mid-altitude sample.

play a major role in determining who does and who does not adopt hybrid seeds. Predicted poverty and food insecurity are lower for the hybrid user group.

Columns 4 and 5 of Table 2 report characteristics of farms in the higher altitude transitional zone in both western and central Kenya. Consistent with Figure 1, hybrid and fertilizer use are uniformly high and maize yields top those of even hybrid users in the mid-altitude zones, as would be expected given that existing hybrids are better adapted to this agro-ecological zone. Incomes are higher and poverty indicators are lower than in the mid-altitude zone, especially for farmers in the central region who are able to grow coffee and other cash crops and who correspondingly get much less of their agricultural income from maize. While farmers in the transitional zone are largely better-off than those in the mid-altitude zones, material well-being is still low relative to global standards.

The lower portion of Table 2 reports compliance with the seed treatment, defined as the percentage of farmers who responded to the informational treatment and purchased a Western Seed variety during the midline and endline seasons. Regression estimates of compliance are reported in Appendix D. In 2014, compliance by the treatment group was modest on average in each zone, ranging from 5%-25%. Field reports indicated a number of factors that conspired to lower uptake that year.[15] These challenges motivated a seed delivery program in 2015. Adoption in 2015 was 5-6 percentage points greater than adoption in 2014 for each of three agro-ecological zones. The different intensity of treatment in 2014 and 2015 of course could lead different types of households to adopt Western Seed hybrids in those years; because of this, in Section 5 we report separate estimates for the endline year.

Finally, our ability to make inferences from the RCT critically depends on the assumption that the random assignment of households to treatment groups is uncorrelated with household characteristics, both observable and unobservable. To shed light on the validity of that assumption, Appendix E presents balance tables for observable characteristics for our sample. In general, baseline differences between treatment groups are not large in magnitude relative to average baseline levels in the control group. Balance on observables gives us confidence that omitting these variables from our estimation will not bias our treatment effect estimates. However, our treatment effect estimates may be biased if a baseline variable

---

[15] Adoption in 2014 was lower than anticipated due to a number of factors, including the former parastatal Kenya Seed subsidizing its seeds, offering added incentives to agro-dealers to sell Kenya Seed. In addition, Western Seed faced challenges in expanding their seed promotion to new regions like central Kenya.

Table 2: Baseline characteristics and compliance.

	Mid-Altitude			**Transitional**	
	All	*Non-Users*	*Users*	*Western*	*Central*
% Main Maize Seasons used Hybrids	26%	11%	96%	84%	83%
% Main Maize Seasons used Fertilizer	26%	20%	56%	81%	92%
Dry Maize Yield (Kg/Acre)	234	211	342	553	428
Acres Farmed in Total	1.65	1.58	1.97	1.92	1.29
Acres Planted to Maize	1.32	1.29	1.44	1.32	0.76
Income Per Capita (100 Kenya Sh.)	228	210	305	351	507
% Net Income from Ag	34%	37%	22%	36%	69%
% Gross Ag Income from Maize	43%	42%	44%	39%	22%
Poverty Probability	32%	33%	26%	32%	13%
% Food Insecure	66%	70%	47%	63%	44%
% Credit Constrained	35%	34%	39%	32%	22%
Compliance: % Using Western Seed					
2014 (Midline), Treated	16%	12%	33%	25%	5%
2014 (Midline), Control	1%	1%	2%	16%	1%
2015 (Endline), Treated	20%	18%	33%	31%	10%
2015 (Endline), Control	2%	1%	8%	17%	0%
Observations	589	482	104	428	508

Notes: "Non-Users" are households that did not plant hybrids in at least 4 of the 5 pre-study main maize seasons and "Users" are households that planted hybrids in at least 4 of the 5 pre-study main maize seasons.

with modest imbalance across treatment groups is strongly correlated with the dependent variable. Since this is most likely to be true for baseline levels of the dependent variable, we estimate analysis of covariance (ANCOVA) specifications that control for baseline levels of dependent variables. Finally, balance on baseline measures of household characteristics gives us confidence that we also have balance on household characteristics that we cannot observe.

4. Average Effects on Yields by Agro-Ecological Zone

This section uses data from our RCT to test whether Western Seed's adapted improved seed varieties increase farmer yields. To account for differences in socioeconomic status and maize seed markets across agro-ecological zones, we estimate the effects of Western Seed hybrids separately for each zone. We find large impacts in the mid-altitude zone, but little to no impact in the transitional zones of western and central Kenya.

To identify the impact of Western Seed maize hybrids on maize yields of smallholder farmers in our study areas, we estimate average treatment effects of the random variation in access to Western Seed varieties. In the mid-altitude and transitional zones, following McKenzie (2012) we estimate average treatment effects on maize yields, y_{ivspt}, using a pooled analysis of covariance intention to treat (ITT) specification.[16] The specification for farm i in village v, site s, matched pair p, and time period t is:

$$y_{ivspt} = \beta_0 + \beta_1 y^0_{ivsp} + \delta_1 I^W_{sp} + \delta_2 I^F_{ivsp} + \delta_3 \left[I^W_{sp} \times I^F_{ivsp}\right] + \left[\mu_p + \varepsilon_{ivspt}\right], t = 1, 2 \qquad (1)$$

where y^0_{ivsp} is baseline maize yields; I^W_{sp} and I^F_{ivsp} are binary indicators for assignment to the Western Seed and fertilizer treatments respectively; μ_p is a matched pair fixed effect included to account for stratification of the seed treatment following Bruhn and McKenzie (2009); and ε_{ivspt} is an error term. In the central region, where there was no fertilizer lottery, we estimate Eq. (1) after eliminating the terms involving I^F_{ivsp}.

In the main body of the paper, we report results for impacts on physical maize yields, measured in kilograms per acre (kg/ac). Appendix G reports the results of estimating Eq. (1) for net maize income per acre instead of physical yields.[17] The results largely parallel the yield results, and will be discussed later. To lessen the role of outliers, we transform yields using the inverse hyperbolic sine transformation (IHST). The lower half of the each table reports the implied percentage effects of the different treatments following the method of Bellemare and Wichman (2020).[18]

We cluster our standard errors at the level of treatment stratification, following Cameron and Miller (2015) and de Chaisemartin and Ramirez-Cuellar (2020). Our experimental design, however, has two treatments, each stratified at a different level, with fertilizer stratified by village and seed stratified by site. But clustering by site would give only 18 clusters in our full sample, and fewer than 20 clusters may lead to biased standard errors (Cameron and Miller, 2015; de Chaisemartin and Ramirez-Cuellar, 2020). Bias from having few clusters would be exacerbated by the fact that there are fewer than 18 sites in each of our sub-samples: the mid-altitude zone (7 sites), the western transitional zone (5 sites), and the central transitional zone (6 sites). Therefore, we cluster standard errors at the level of stratification for the fertilizer treatment, the village. In Appendix G, we report results using

---

[16] Our preferred specification estimates a single set of treatment effects by pooling observations from both of the post-treatment years of observations. When separating observations by these two post-treatment years, treatment effect estimates are qualitatively similar to the estimates from our preferred specification, both for the analysis in this section as well as the heterogeneity analysis in the subsequent section.

[17] Net maize income is defined as the value of production minus the cost of seeds and fertilizers.

[18] For example for the seed treatment, the estimated percentage change is calculated as $\exp(\hat{\delta_1} - 0.5\hat{\sigma}^2_{\hat{\delta}_1}) - 1$, where $\hat{\sigma}^2_{\hat{\delta}_1}$ is the estimated variance of $\hat{\delta}_1$. As shown by Goldberger (1968), this estimated percentage change is upwardly biased, but this bias vanishes asymptotically. While this point appears to have been overlooked by Bellemare and Wichman (2020), it does raise issues about the reliability of this measure in small samples.

an alternative approach recommended for experimental settings with few clusters, randomization inference, which is based on design uncertainty in our experimental design rather than sampling uncertainty. Randomization inference supports the conclusions that we draw from our clustered standard errors in the main text of the paper.

Table 3 reports the estimated coefficients for Eq. (1) by agro-ecological zone. For the mid-altitude zone, we estimate that the seed treatment increased average yields by 25%. Given that these are intention to treat estimates (with a compliance rate of 16% as reported in Table A2), the actual yield increases experienced by those who adopted Western Seed because of the seed treatment (impact of the treatment on the treated) is approximately six-times greater, indicating substantial yield gains among adopters of Western Seed. That said, the post-treatment difference between agronomically-potential yields in Figure 3 and average realized yields in our sample only closes from 90% to about 80% with this doubling of yields.

Given that the mid-altitude area is one with relatively low hybrid use, these average impacts of Western Seed likely come from two sources: First generation adoption effects from the intervention inducing households to switch from local varieties to hybrid varieties, and second generation adoption effects from the intervention inducing households to switch from other hybrid varieties to Western Seed locally adapted hybrid varieties. In the next section, we explore this issue further to distinguish the magnitudes of these two effects.

Table 3: Effects on maize yield (IHST of kg/ac).

	Mid-Altitude	**Transitional**	
	All Farms	*Western*	*Central*
Seed Treatment, $\hat{\delta}_1$	0.23*	-0.17	0.10
	(0.13)	(0.12)	(0.15)
Fertilizer Treatment, $\hat{\delta}_2$	0.23**	0.11	–
	(0.11)	(0.09)	–
Seed*Fertilizer Treatment, $\hat{\delta}_3$	-0.23*	0.20	–
	(0.14)	(0.12)	–
Baseline Yield, $\hat{\beta}_1$	0.11***	0.15***	0.20***
	(0.03)	(0.02)	(0.05)
Percent effects			
Seed Treatment	25%	-17%	9%
Fertilizer Treatment	26%	11%	–
Control mean	5.97	6.86	6.27
Observations	1178	856	1016
R-squared	0.07	0.06	0.11

Pair indicator variables included as controls.
Standard errors clustered by village.
* = 10% significance, ** = 5%, *** = 1%

In addition to the primary seed treatment effect on yields in the mid-altitude zone, the other striking result in Table 3 is the negative interaction effect between the seed and fertilizer treatments. As discussed in Sub-Section 3.1 above, in-kind grants of fifty kilograms of fertilizer were made to randomly selected study households in order to see if relaxing constraints to fertilizer acquisition might boost returns to the seed treatment. If access to fertilizer had proven not to be a constraint, we might have expected to see the impact of the fertilizer grant, both alone and in combination with the seed treatment, to be zero. Instead, Table 3 shows a positive effect of the direct impact and a negative impact of the interaction effect. Given how we have defined our treatment assignment variables, the impact of winning the fertilizer lottery in seed treatment zones is the sum of the coefficient on fertilizer plus the coefficient on the interaction term. The estimated coefficients imply a point estimate of nearly zero additional impact for lottery winners versus lottery losers in seed treatment areas. This stands in stark contrast to our expectation that the coefficient on the interaction terms would be positive given the expectation that improved varieties are more fertilizer responsive than local, retained varieties.

In order to better understand this unexpected outcome, we look more closely at the impact of the different treatments on fertilizer use in Appendix H. There we find substantially greater midline leakage (or sharing) of fertilizer from lottery winners in seed treatment communities than in non-seed treatment communities. During the midline season, the fertilizer use differential between winners and losers in seed control communities is 25 kg, whereas it is only 10 kg in the seed treatment communities. This apparently differential leakage of fertilizer from lottery winners suggests there were greater rates of social taxation of "windfall fertilizer" in seed treatment sites. Given this configuration, it is unsurprising that households that benefited less from the fertilizer treatment in seed treatment communities than in control communities. As detailed in Appendix H, this differential rate of taxation likely explains an important fraction of this puzzle. As further discussed in the appendix, other explanations are possible (*e.g.,* shared fertilizers went disproportionately to those who knew how to use them effectively), but our data do not allow us to test these other explanations.

Finally, while we do see significant impacts of the seed treatment in the mid-altitude zone, in the transitional zone the estimated treatment effects on yields are relatively small and are not significantly different from zero at even the ten percent level. Recall that hybrid use was already quite high in this zone prior to the study, and that net compliance was also quite low (see Table 2).

5. Heterogeneous Effects on Yields by Past Hybrid Use

As seen in Table 3, we find that the introduction of Western Seed varieties impacts maize yields in the mid-altitude zone. Given that prior adoption of hybrids in this area was modest, the question remains whether the observed impacts reflect the fact that Western Seed varieties outperform local varieties as well as other commercially available hybrids (largely bred for higher altitude maize-growing areas). The theoretical model developed in Section 2 is silent on the question as to which group will benefit most from a well-adapted improved seed: those who make the switch from local, unimproved seed or those that switch from less well-adapted improved seed. Using stylized assumptions, the model does show that the impacts could be quite different for these groups and therefore that the just estimated average treatment effects may obscure important differences across farmer types. To gain empirical purchase on this variation, this section splits our mid-altitude sample between households that consistently used hybrids prior to the study period and those that did not (see Table 2). In terms of the theoretical model summarized in Figure 2, this section compares impacts of Group 1 farmers with impacts on Group 2.

Table 4 displays the ITT estimates for Eq. (1) for the hybrid users and hybrid non-users sub-samples. The bottom panel of Table 4 reports that the estimated impacts of the seed treatment are 21% for the non-user group and 47% for the user group, with only the latter estimate attaining conventional levels of statistical significance. As shown in Table 2, net compliance for the users group was 31% in the midline and fell to 25% at the time of the endline (with the decline driven by a substantial uptick in Western Seed use by the control group). For the non-users group, net compliance was only 11% in midline, and rose to 17% in the endline following the ancillary seed delivery intervention. This lower level of compliance explains at least in part the smaller magnitude a of the ITT estimate for the non-users group. For the users group, the estimated 47% ITT impact translates into a robust impact for those who actually adopted the seeds.

Given the compliance issues that especially surrounded the midline year, the research team implemented an ancillary seed delivery service, as described in Sub-Section 3.2 above. While the delivery was less effective than anticipated, it did boost compliance among the non-users group.[19] The second and fourth columns of Table 4 thus present separate estimates

---

[19] Phone orders for the seed were quite high, with 55% of treatment farmers ordering seed. Unfortunately, at time of delivery, only 16% of treatment farmers bought seed via the delivery intervention. Many farmers

Table 4: Heterogeneous impact of locally adapted improved varieties in the mid-altitude zone (IHST of kg/ac).

	Mid-Altitude Zone			
	Hybrid Non-Users		*Hybrid Users*	
	Pooled	Endline	Pooled	Endline
Seed Treatment, $\hat{\delta}_1$	0.20	0.26*	0.40**	0.78***
	(0.15)	(0.15)	(0.16)	(0.21)
Fertilizer Treatment, $\hat{\delta}_2$	0.22	0.18	0.35**	0.54**
	(0.13)	(0.14)	(0.16)	(0.21)
Seed*Fertilizer Treatment, $\hat{\delta}_3$	-0.21	-0.21	-0.25	-0.44*
	(0.17)	(0.18)	(0.24)	(0.26)
Baseline Yield, $\hat{\beta}_1$	0.12***	0.09**	0.02	-0.02
	(0.03)	(0.04)	(0.05)	(0.05)
Percent effect				
- Seed Treatment	21%	29%	47%	113%
- Fertilizer Treatment	23%	18%	39%	68%
Control mean	5.92	6.03	6.13	5.98
Observations	964	482	208	104
R-squared	0.06	0.03	0.19	0.22

Pair indicator variables included as controls. Standard errors clustered by village.
Users planted hybrids in 80-100% of main seasons (2007-2012).
* = 10% significance, ** = 5%, *** = 1%

for the endline year data only. As can be seen, for the non-users group, the estimated impact rises and becomes statistically significant and indicates an ITT treatment effect of a 29% yield increase. The estimated impact for the hybrid users group retains its significance and jumps substantially to an implied 113% yield increase.

While these results indicate large and positive average effects of the Western Seed treatment, the question remains how large the yield gains are for farmers who actually adopt Western Seed. A standard approach to answer this question would be to instrument for Western Seed adoption with assignment to treatment to estimate a local average treatment effect. However, the low adoption rates in our sample make treatment assignment a weak instrument for adoption, as can be seen by the low F-statistics in Table A2. Thus, instrumental variables estimation of the local average treatment effect likely would yield biased point estimates and standard errors. Instead, we approximate the local average treatment effect by adjusting the percent effect implied by our ITT estimates by the net compliance rates for each group.[20] This approach suggests that compliers with the Western Seed treat-

---

proved unwilling or unable to pay for the seed despite the fact that it had been made clear that the seeds were being sold at the market price and not given away for free.

[20] This admittedly *ad hoc* approach is intended to approximate the yield gains experienced by adopters and allow the reader to obtain a sense of the impact of the treatment on the very large yield gap observed at baseline in the data.

ment increased yields by 150% among non-users (170% if we use the estimates based on endline impacts only) and 168% among users (452% endline only). To put these numbers in context, note that baseline yields for non-user and user groups were 211 and 342 kg/acre, respectively (Table 2). Compared to these numbers, the estimated percentage changes imply post-treatment yields for compliers in the non-user groups of 528-570 kg/acre, and 917-1888 kg/acre for the user group.

While these yield increases are large, a substantial gap remains between the yields realized by farmers in our sample and the potential yields shown in Figure 3. While it is a truism that real farmers never obtain the yields obtained on carefully tended pilot and experiment station plots, the magnitude of the yield gap in part reflects the fact that treated study farmers' use relatively low levels of fertilizer and only partially rely on improved seed.[21] Credit, risk, and perhaps behavioral constraints likely help explain these patterns of modest input use and the residual yield gap.

While these results on physical yields show gains, it is important to ask if yield gains of these magnitudes are profitable. To gauge the economic impact of the intervention, we use the survey data to calculate net revenue (the value of production less seed and fertilizer costs) for each farmer in the sample.[22] This net revenue measure does not account for the value of land nor labor in maize production due to our lack of detailed data on these factors of production. It is therefore a measure of returns to land and labor rather than a measure of economic profits.

Using this net revenue per acre measure as the dependent variable, we re-estimate Eq. (1) to obtain ITT estimates of the impact of the treatments on farm household earnings. Table 5 estimates heterogeneous impacts of treatment in the mid-altitude zone by baseline hybrid use.[23] The bottom panel of Table 5 reports that the estimated impacts of the seed treatment are 20% for the non-user group and 76% for the user group, with only the latter estimate attaining conventional levels of statistical significance. These figures rise to 33% and 219% if we consider only the endline data, although the former number is based on

---

[21] Part of the standard explanation is that farmers are interested in income, not maximizing yields

[22] Because of high variability in prices reported by farmers, we value each unit of maize production at the average unit value of maize sold in the sample in a given season. Costs of maize production include costs of seeds and fertilizers; for each seed variety and fertilizer type, we value each unit at the average unit value purchased in the sample in a given season. We value fertilizer distributed as part of the study at the average unit value for the most comparable fertilizer in our sample (NPK 23:23:0).

[23] Appendix Table A7 estimates average effects by agro-ecological zone. Unsurprisingly, given the findings on yield impacts at this level, this table shows a positive (but statistically insignificant) gain from the treatment in the mid-altitude zone and no economic gain from the treatment in the transitional zones.

Table 5: Heterogeneous impacts on net revenue per acre in the mid-altitude zone (IHST of ksh/ac).

	Mid-Altitude Zone			
	Hybrid Non-Users		*Hybrid Users*	
	Pooled	Endline	Pooled	Endline
Seed Treatment, $\hat{\delta}_1$	0.22	0.31	0.61**	1.32**
	(0.26)	(0.25)	(0.29)	(0.57)
Fertilizer Treatment, $\hat{\delta}_2$	0.04	0.20	0.52	0.84
	(0.28)	(0.27)	(0.32)	(0.58)
Seed*Fertilizer Treatment, $\hat{\delta}_3$	-0.07	-0.17	-1.10*	-1.51
	(0.33)	(0.34)	(0.64)	(0.98)
Baseline Economic Yield, $\hat{\beta}_1$	0.07*	0.00	-0.02	-0.06
	(0.04)	(0.02)	(0.04)	(0.05)
Percent effect				
- Seed Treatment	20%	33%	76%	219%
- Fertilizer Treatment	0%	18%	61%	95%
Control mean	8.90	9.05	9.11	8.80
Observations	964	482	208	104
R-squared	0.02	0.01	0.10	0.11

Pair indicator variables included as controls. Standard errors clustered by village.
Users planted hybrids in 80-100% of main seasons (2007-2012).
* = 10% significance, ** = 5%, *** = 1%

a statistically insignificant coefficient. That said, in line with the yield results, these ITT point estimates signal that both groups of farmers likely benefitted financially from the seed treatment. These financial figures place the impacts squarely in Case 1 of Table 1 above, with both an intercept and a slope effect such that Western Seed varieties benefits both farmers previously using retained varieties and farmers previously using non-locally adapted hybrid varieties.

6. Conclusion

While the average Kenyan farmer deviates from the sub-Saharan African pattern of low hybrid maize adoption, the mid-altitude zone of Kenya more closely resembles the rest of the continent: persistently modest hybrid use over the last 50 years. Moreover, prior to the experiment reported here, the only hybrid varieties available to mid-altitude farmers were those adapted to Kenya's higher altitude regions. One simple explanation for this pattern is that it is not possible to breed high productivity maize varieties for the mid-altitude environment. Another explanation is that while such innovation is possible, a confluence of demand and supply patterns may create an equilibrium in which no firm innovates these varieties given the cost and other constraints that they face. From this perspective, large-scale seed companies do not supply well-adapted hybrid varieties to small markets. Neither do local seed companies, who may enjoy informational cost-advantages for innovation, but are constrained by their lack of capital. When varieties adapted to the local, niche agro-ecology are not available on the market, our theoretical model of seed variety adoption indicates that adoption of the available (non-adapted) hybrids will be low and restricted to better-resourced and more knowledgable farmers within the niche market. Absent locally adapted hybrids, farmer productivity and incomes stagnate, especially among the poorest farmers. Pre-intervention yields appear to be only about 10% of what appears to be agronomically possible.

Against this backdrop, an external infusion of impact investor capital to a local seed company based in Kenya's mid-altitude zone allowed us to explore the impacts of disrupting this seed market equilibrium. By 2010, Western Seed had developed and registered hybrid maize varieties—building on publicly-provided parent seed lines from CIMMYT—and were supplying high-performing locally adapted seed varieties to the mid-altitude niche market. The impact investment capital allowed the company to rapidly expand its seed breeding and production capacity and expand its marketing to the mid-altitude and the (higher altitude) transitional zone. In collaboration with Western Seed, we established a three-year RCT to study the impact of offering the new seed varieties.

The key findings from the empirical analysis are that the new varieties offered substantial benefits to farmers located in the mid-altitude region. On average, the ITT estimate of the yield impact implies a 25% increase for farmers in the mid-altitude treatment group. Given the modest compliance rates, the actual impact on adopters was substantially higher. In the higher-altitude zones where Western Seed also expanded, the yield benefits are both statistically negligible, consistent with the expectation that the seed sector has already adapted varieties for these larger and more lucrative zones.

Consistent with our theoretical model, the data reveal that prior to the experimental introduction of Western Seed hybrids, farmers in the mid-altitude zone could be divided into two groups: a larger group (82% of the sampled farmers) who almost never use improved seeds nor apply chemical fertilizer, and a smaller group (18%) who almost always use improved seeds and fertilizer. While our statistical power suffers as we sub-divide the mid-altitude farmers into these two groups, we estimate that the yield impacts on farmers already using hybrids are actually larger than those on the farmers who switched from local, unimproved seed varieties. In the last year of the RCT, ITT estimates indicate that the physical productivity increases were 29% for those who rarely sued hybrids, and 113% for prior hybrid adopters who are also able to apply fertilizers to their maize. While large, these figures are well within officially registered yield potential of these varieties in the mid-altitude region, even after accounting for experimental compliance rates that range between about 15 and 30%. Estimates of the net-revenue gains from the seed treatment are positive and broadly similar to the reported yield increases.

Stepping back, these results indicate that productivity- and income-enhancing opportunities were being left on the table by the prevailing seed system. While our data do not allow us to test the performance of different seed systems *per se*, the patterns that we observe are consistent with a theoretical model in which small agro-ecological niches will remain underserved absent the multiple partnerships that allowed Western Seed to expand (specifically public investments in foundation seed and impact investment capital). Our results cautiously suggest that such a hybrid system has much to offer other areas of sub-Saharan Africa where average cereal yields lag far behind the technological frontier, as they do in Kenya's mid-altitude region.[24] These observations do not say that the other constraints discussed in the introduction do not matter. Indeed, even the better resourced treated farmers in our study, who regularly buy some fertilizer, are getting only half of the yield that is technologically available in this region. However, better adapted and more profitable improved varieties do have consequential impacts for farmers, and might ultimately alter their ability to overcome these other constraints.

Entry of new actors into niche breeding may transform innovation markets for agricultural technologies in regions like mid-altitude Kenya to more closely resemble the present in regions like transitional and highland Kenya, where many firms compete in hybrid development. Such a change would shift research priorities toward studying the implications of competition and product differentiation between firms on the productivity and welfare of agricultural households in these environments. From this perspective, our finding that locally adapted varieties offer important benefits to both poor and better-resourced farmers indicates that such adaptive breeding, whatever its source, can offer substantial economic and social benefits.

---

[24] Our model in Appendix C assumes that local companies have lower fixed costs of hybrid development due to greater knowledge of the local growing conditions. While we believe this assumption approximates the case of hybrid development in Kenya in the early 2000s, recent advances in genetics and big data analysis may tilt the cost advantage in local adaptation in the favor of large multinational firms that can realize economies of scope from drawing on genetic markers and data from agronomic test trials in agro-ecological environments around the world. Time will tell whether these new advances suffice to incentivize niche adaptive breeding for the large firms that dominate the global hybrid maize market.

Acknowledgements

We are grateful for comments from Brian Dillon, Jim Gaffney, GianCarlo Moschini, Wendong Zhang, and seminar participants at PacDev 2021, MWIEDC 2021, Iowa State University, the University of California–Davis, and the University of Wisconsin–Madison. We thank our survey respondents for their enthusiasm, generosity, and patience during the course of this research. We also thank the management of Western Seed for their patient understanding and willingness to cooperate with the rigors of a randomized control trial built around their market expansion.

Funding: This work was supported by Acumen, a non-profit impact investment firm; the American people through the United States Agency for International Development Cooperative Agreement No. AID-OAA-L-12-00001 with the BASIS Feed the Future Innovation Lab; and the Agricultural Technology Adoption Initiative (ATAI) administered by JPAL at MIT and the Bill and Melinda Gates Foundation. The contents are the responsibility of the authors and do not necessarily reflect the views of Acumen, USAID, the US Government, or other funders. Declarations of interest: none.

References

Abay, K., Blalock, G., Berhane, G., 2017. Locus of control and technology adoption indeveloping country agriculture: Evidence from ethiopia. Journal of Economic Behavior and Organization 143, 98–115.

Bellemare, M.F., Wichman, C.J., 2020. Elasticities and the Inverse Hyperbolic Sine Transformation. Oxford Bulletin of Economics and Statistics 82, 50–61. URL: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/obes.12325`.

Besley, T., Case, A., 1993. Modeling Technology Adoption in Developing Countries. American Economic Review 83, 396–402. URL: `https://ideas.repec.org/a/aea/aecrev/v83y1993i2p396-402.html`.

Bliss, C., Stern, N., 1978. Productivity, wages and nutrition: Part i: the theory. Journal of Development Economics 5, 331–362.

Bold, T., Kaizzi, K.C., Svensson, J., Yanagizawa-Drott, D., 2017. Lemon Technologies and Adoption: Measurement, Theory and Evidence from Agricultural Markets in Uganda*. The Quarterly Journal of Economics 132, 1055–1100. URL: `https://doi.org/10.1093/qje/qjx009`, doi:10.1093/qje/qjx009, `arXiv:https://academic.oup.com/qje/article-pdf/132/3/1055/30646731/qjx009.pdf`.

Bruhn, M., McKenzie, D., 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. American Economic Journal: Applied Economics 1, 200–232.

Cameron, A.C., Miller, D., 2015. A practitioner's guide to cluster-robust inference. Journal of Human Resources 50, 317–372.

Carter, M.R., Laajaj, R., Yang, D., 2021. Temporary Subsidies and Green Revolution Technology Adoption by Farmers and Their Social Networks. American Economic Journal: Applied Economics 13, 206–229.

de Chaisemartin, C., Ramirez-Cuellar, J., 2020. At What Level Should One Cluster Standard Errors in Paired Experiments, and in Stratified Experiments with Small Strata? Working Paper 27609. National Bureau of Economic Research. URL: `http://www.nber.org/papers/w27609`, doi:10.3386/w27609.

Conley, T.G., Udry, C.R., 2010. Learning about a new technology: Pineapple in ghana. American Economic Review 100, 35–69. URL: `https://www.aeaweb.org/articles?id=10.1257/aer.100.1.35`, doi:10.1257/aer.100.1.35.

Dercon, S., Christiaensen, L., 2011. Consumption risk, technology adoption and poverty traps: evidence from ethiopia. Journal of Development Economics 96, 159–173. Working paper.

Duflo, E., Kremer, M., Robinson, J., 2011. Nudging farmers to use fertilizer: Theory and experimental evidence from kenya. American Economic Review 101, 2350–2390.

Evenson, R.E., Gollin, D., 2003. Assessing the impact of the Green Revolution, 1960 to 2000. Science 300, 758–762. doi:10.1126/science.1078710.

Foster, A.D., Rosenzweig, M.R., 1995. Learning by doing and learning from others: Human capital and technical change in agriculture. Journal of Political Economy 103, 1176–1209. URL: `https://doi.org/10.1086/601447`, doi:10.1086/601447, `arXiv:https://doi.org/10.1086/601447`.

Gerhart, J., 1975. The diffusion of hybrid maize in Western Kenya—Abridged by CIMMYT. Technical Report. Centro Internacional de Mejoramiento de Maiz y Trigo. Mexico City.

Goldberger, A.S., 1968. The interpretation and estimation of cobb-douglas functions. Econometrica 36, 464–472.

Griliches, Z., 1960. Hybrid Corn and the Economics of Innovation. Science 132, 275–80. URL: `http://www.ncbi.nlm.nih.gov/pubmed/17756993`, doi:10.1126/science.132.3422.275.

Hassan, R.M., 1998. Maize Technology Development and Transfer: A GIS Application for Research Planning In Kenya. CAB International, Wallingford and New York.

Jayne, T.S., Mason, N.M., Burke, W.J., Ariga, J., 2018. Review: Taking stock of africa's second-generation agricultural input subsidy programs. Food Policy 75, 1–14. URL: `https://www.sciencedirect.com/science/article/pii/S0306919217308618`, doi:https://doi.org/10.1016/j.foodpol.2018.01.003.

Kantor, S., Whalley, A., 2019. Research Proximity and Productivity: Long-Term Evidence from Agriculture. Journal of Political Economy 127.

Karlan, D., Osei, R., Osei-Akoto, I., Udry, C., 2014. Agricultural decisions after relaxing credit and risk constraints. Quarterly Journal of Economics , 597–652.

KEPHIS, 2018. National Crop Variety List - Kenya. Technical Report. Kenya Plant Health Inspectorate Service. Nairobi.

Laajaj, R., Macours, K., forthcoming. Measuring skills in developing countries. Journal of Human Resources .

Langyintuo, A.S., Mwangi, W., Diallo, A.O., MacRobert, J., Dixon, J., Bänziger, M., 2010. Challenges of the maize seed industry in eastern and southern Africa: A compelling case for private-public intervention to promote growth. Food Policy 35, 323–331. URL: `http://dx.doi.org/10.1016/j.foodpol.2010.01.005`, doi:10.1016/j.foodpol.2010.01.005.

Marenya, P., Barrett, C.B., 2009. State-conditional fertilizer yield response on western kenyan farms. Amer. J. Agr. Econ. 91, 991–1006.

McKenzie, D., 2012. Beyond baseline and follow-up: The case for more T in experiments. Journal of Development Economics 99, 210–221.

Munshi, K., 2004. Social learning in a heterogeneous population: technology diffusion in the indian green revolution. Journal of Development Economics 73, 185–213. URL: `https://www.sciencedirect.com/science/article/pii/S0304387803001342`, doi:https://doi.org/10.1016/j.jdeveco.2003.03.003.

Partners, P.C., no date. Impact investment Understanding financial and social impact of investments in East African agricultural businesses. Technical Report. The Gatsby Charitable Foundation.

Suri, T., 2011. Selection and comparative advantage in technology adoption. Econometrica 79, 159–209.

TAMPA2, 2004. Tegemeo Agricultural Monitoring and Policy Analysis Project. Technical Report. Egerton University - Tegemeo Institute of Agricultural Policy and Development and Michigan State University. Nairobi. URL: `https://www.tegemeo.org/images/downloads/data/2004-Survey-Documentation.pdf`.

TAPRA, 2010. Tegemeo Agricultural Policy Research Analysis Project. Technical Report. Egerton University - Tegemeo Institute of Agricultural Policy and Development and Michigan State University. Nairobi. URL: `https://www.tegemeo.org/images/downloads/data/2010_SurveyDocumentation.pdf`.

Tjernstrom, E., 2017. Learning from others in heterogeneous environments.

Appendix A: Linear Returns to Fertilizer

As Section 2 describes, standard agronomic practice recommends that farmers who cannot apply the optimal amount of fertilizer to their entire crop, should apply the optimum rate to a portion of their crop rather than applying fertilizer at a diluted rate to the entire crop. Using a simple economic model that has roots in the efficiency wage literature's portrayal of the response of human work capacity to nutrition (see Bliss and Stern, 1978), this appendix shows that optimized returns to fertilizer will be linear under profit maximization whenever the farmer is constrained in her purchase of fertilizer.

Figure A1 graphs the additional production per-hectare (q) that occurs as fertilizer intensity (f) increases under the assumption that marginal returns to fertilizer exhibit increasing returns over a range and then diminishing returns thereafter. Without fertilizer, farmers receive a fixed amount α_0, and $q(f)$ is the additional output received in addition to the base, no fertilizer amount. In the figure, the slope of the ray from origin is the average product of fertilizer (q/f), or the bang for the buck spent on fertilizer. As can be seen by visual inspection, f^* is the fertilizer intensity that maximizes additional production per-unit fertilizer given the S-shaped production function. No other intensity will give more. We denote the slope of the ray that intersects the function q at input level f^* as α_1.

Consider the case where $f^* = 100$ kg/ha and a farmer with one hectare of maize has only 50 kg of fertilizer, then she will maximize output and income by concentrating the 50 kg on 0.5 hectare and putting zero fertilizer on the rest. As can easily be confirmed visually, alternative allocations (*e.g.*, 50 kg/hectare on her entire maize plot, yielding a fertilizer intensity of $\frac{f^*}{2}$) will yield less output for the same input expenditure because it fails to fully exploit the increasing returns portion of the returns to fertilizer function, $q(f)$.[25] That is, for this case of constrained access to fertilizer:

$$0.5q(0) + 0.5q(f^*) > 0.5q(\varepsilon) + 0.5q(f^* - \varepsilon), \forall\varepsilon.$$

For this case, when fertilizer is optimally applied, marginal returns to additional units of fertilizer will always be a constant α_1, not because marginal returns do not change, but

---

[25] This argument is exactly identical to the initial contributions to the efficiency wage literature in which an employer (who is indirectly buying nutrition by paying workers a wage) will never pay a worker less than the efficiency wage because worker productivity falls off more quickly than cost when descending the increasing returns portion of the efficiency wage curve.

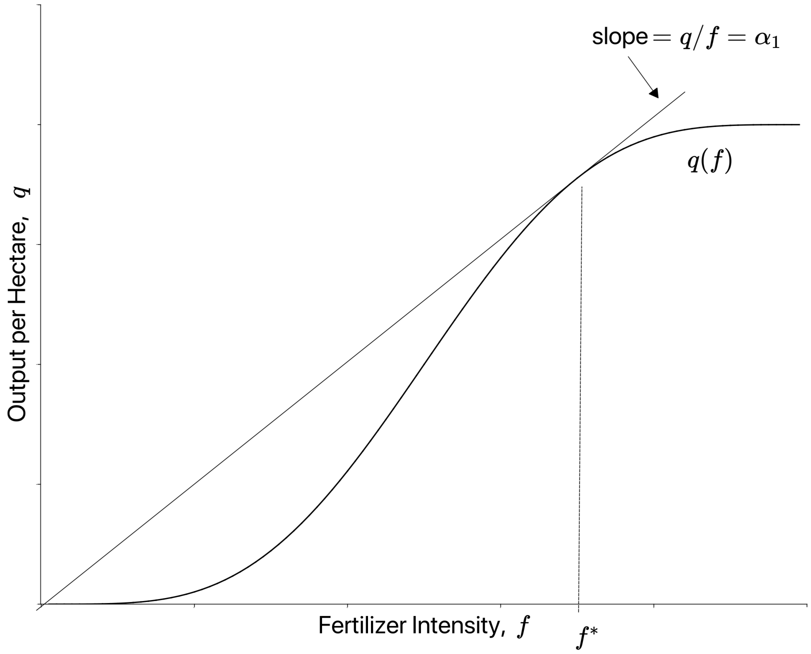


Figure A1: Returns to fertilizer.

because the farmer optimally adjusts fertilizer application to obtain the constant, maximal return. Once the farmer can apply fertilizer at intensity f^* on all her land, further use of fertilizer will face diminishing returns as beyond that level, the function $q(f)$ is strictly concave.

This intuition can be captured by the following profit maximization problem:

$$\underset{F,\,T_f}{Max} \quad p_m Q - p_f \bar{F} \tag{2}$$

$$\begin{aligned} &subject\ to: \\ &\qquad Q = T_f q(F/T_f) + \left(\bar{T} - T_f\right) q(0) \\ &\qquad p_f F \leq K \\ &\qquad T_f < \bar{T} \end{aligned} \tag{3}$$

where p_f is the price of fertilizer and K is the working capital available to purchase fertilizer. Note that this simple specification allows any number of possible outcomes, including zero use of fertilizer ($F, T_f = 0$), which will occur when $\alpha_1 p_m < p_f$; and, use of fertilizer in excess of f^* ($T_f = \bar{T}$ and $F > f^* \bar{T}$), which will happen when $K > p_f \bar{T} f^*$ and $\alpha_1 p_m > p_f$. As discussed in Sub-Section 2.1, the relevant case for low income small scale farmers, who cannot borrow against their future income to buy fertilizer, is that $K < p_f \bar{T} f^*$. For these farmers, the *optimized* production function they face can thus be written as $\alpha_0 + \alpha_1 f$, with constant marginal returns to fertilizer. As further analyzed in Sub-Section 2.1, variety choice and the decision whether or not to use any fertilizer will be based on those linear returns for liquidity constrained farmers.

Appendix B: Modeling the Demand for Improved Seed Varieties

We assume that the farming input and consumption choices of a household with $\bar{H}$ units of land, z_0 units of cash-on-hand and technology pessimism parameter λ are guided by the following two-period expected utility maximization model:

$$\underset{\tilde{H}^v, I^v, f^v, S_1}{Max} \quad u(c_1) + \beta E_\theta \left[u(c_2) \right] \tag{4}$$

$$\begin{aligned} & \textit{subject to}: \\ & c_1 \leq z_0 - \sum_{v=r,n,a} H^v \left(s^v + p^f f^v \right) - S_1 \\ & c_2 \leq y^{noag} + (1 + r^i) S_1 + \sum_{v=r,n,a} H^v \theta \lambda^v p_m \left(\alpha_0^v + \alpha_1^v f^v \right) \\ & \sum_{v=r,n,a} H^v \leq \bar{H} \\ & I^f (H^v f^v - \underline{\mathrm{F}}) = 0, \forall v \\ & I^v (\tilde{H}^v - \underline{\mathrm{H}}^v) \geq 0, \forall v \\ & f^v \geq 0, \forall v \end{aligned}$$

where $u(.)$ is a concave utility function, S_1 is savings carried forward from period 1 to 2 at interest rate r^i, and $\beta < 1$ is the per-period discount factor. We assume that the discount rate underlying β is strictly greater than r^i such that individuals will only save to manage period 2 risk. The indicator variables, I^f and I^v, take on the value of 1 when, respectively, fertilizer and area devoted to variety v become strictly greater than 0. These constraints allow us to impose the restrictions that amounts of improved seed and fertilizer must at a minimum be one bag.[26] λ^v is the technology pessimism parameter that is less than or equal to 1 for improved varieties, and equals 1 for retained local seed varieties. The stochastic

---

[26] In principal, this seed bag integer problem should continue, but as our analysis only concerns the adoption decision, we will ignore that aspect of the problem.

Table A1: Parameter values for numerical analysis.

Technology Parameters									
Retained		*NLA Hybrid*		*Adapted Hybrid*		*Seeding*	*Min*	*Min*	*Land*
α_0^r	α_1^r	α_0^n	α_1^n	α_0^a	α_1^a	*Rate*	*Seed*	*Fert*	$\bar{H}$
240 kg/acre	5	240kg/acre	20	450 kg/acre	25	10kg/acre	2 kg	10 kg	2 acres
Prices ($US)					**Utility Parameters**				**Production Shock**
Maize, p_m	p_f	s^r	s^r	s^r	*Function*	ρ	β	y^{noag}	$\theta \sim N(1,1)$
0.5/kg	$8/kg	$4	$40	$40	$\frac{1}{(1-\rho)}c^{1-\rho}$	1.6	0.95	$5	$for\ 0 \leq \theta \leq 2$

specification is given in the main body of the paper.

While this model is amenable to general analysis, its implications for purposes of this paper can be most easily gleaned through numerical analysis that we base on parameter values that are faithful to the empirical reality in our study in Kenya. Table A1 lists the parameter values used for the numerical analysis presented in the main body of the paper. The prices reported in the table reflect the values reported by farmers in our survey.[27] Note also that fertilizer becomes a break-even investment for variety v if $\alpha_1^v = 16$. Under the numerical assumptions given in the table, fertilizer is profitable in expectation for both improved varieties, but is not profitable for the local variety, r.

Appendix C: The Supply of Agricultural Innovation to Niche Markets

If at least some farmers would be willing to adopt locally adapted seeds (as would be true in all but Case 4 in Table 1), then what accounts for the 50 years of a low and stagnant hybrid adoption rate in Kenya's mid-altitude zone? One explanation is that improved varieties cannot be successfully adapted to this region (Case 4 in Table 1). If this were the case, we would expect the experimental introduction of putatively locally adapted varieties to show zero, or at least non-profitable impacts. But if such varieties can be developed (Cases 1-3), then what explains their 50-year absence from Kenya's mid-altitude zone? In this section, we consider key features of the seed industry which can explain the failure of seed companies to undertake technologically feasible adaptive breeding for small, agro-ecological niches, especially when populated by farm households of the sort modeled in Sub-Section 2.1. Upon request, the authors can provide a formal model of profit-maximizing innovation that captures the intuitions offered in this appendix.

One factor that influences the economics of seed innovation is how the firm obtains

---

[27] Study respondents receive about $0.26 for a kilogram of maize, pay $0.35 to purchase local seeds for planting and $2.00 per kilogram of nitrogen. Non-Locally adapted hybrids sell for $2.10-$2.40 per kilogram, while the locally adapted variety sold for $1.95 per kilogram.

parent lines for local adaptive breeding. If parent lines are owned and maintained by the firm itself, then the firm pays no royalties for using the parent varieties. Given the high cost of developing and maintaining parent breeding lines, this option is only available to large, multinational firms. Smaller firms can access parent material by purchasing use rights from other private sector firms at a royalty cost per-kilogram of adapted seed produced. In addition, firms may have the option to use without royalty parent seed produced by the public sector.[28] When public sector breeders make parent lines freely available, firms can avoid paying royalty costs for parent material.

Given access to parent seed lines, firms incur a non-trivial fixed cost to breed seed varieties for the local agro-ecology. These costs are related to not only acquiring farm land on which to experiment, but also to acquiring the knowledge about local conditions that limit crop performance. We assume that local firms, which typically emerge from farms already producing in the local area, have a fixed cost advantage over multinational or other non-local firms who need to both acquire land and learn about the particularities of local farm production.

Given these characteristics, the profit-maximizing firm's decision of whether to innovate locally adapted seed varieties depends on the following two constraints.

- *Effective Market Size:* The presence of fixed costs implies that firms will not innovate for a market unless expected sales are large enough to warrant payment of the fixed costs.
- *Capital Constraints:* The adaptation of seeds is a multi-year process of testing and ultimately regulatory approval, meaning firms need to be able to finance up-front royalty and fixed costs.

Given these constraints, a zero innovation equilibrium can emerge for markets that are too small to interest large firms, and capital constraints are too severe for small firms, preventing them from exploiting their fixed cost advantage. In this case, disrupting this equilibrium and providing locally adapted varieties to a niche agro-ecology would require relaxation of financial constraints for local firms and/or partnerships between local firms and public sector breeders who provide royalty-free access to parent seed lines. In the case of Kenya,

---

[28] The public sector has long filled this role, which has, for example, contributed to regional differences in the development and adoption of hybrid maize in the United States (Griliches, 1960; Kantor and Whalley, 2019). In developing countries, public sector investments are supplemented by investments by international organizations through the CGIAR networks, in particular CIMMYT for maize.

Western Seed not only had a partnership with a public source of quality parent line seed, it also received a major infusion of social impact investor capital (not otherwise available on the market) that allowed the firm to rapidly expand its seed multiplication capacity. As the empricial sections of this paper explore, the expansion of this seed company gave us the opportunity to explore whether the seed system in Kenya was indeed leaving money on the table by failing to realize profitable innovation of locally adapted improved maize varieties.

Appendix D: Effects on Inputs by Agro-Ecological Zone

Table A2 displays average effects of treatment on input use by agro-ecological zone. As is the case throughout our empirical analysis, observations are of households observed in all survey rounds and pool both rounds post-treatment.

In the mid-altitude zone, the seed treatment increases adoption of Western Seed Company maize hybrids by 16 percentage points on average across the post-treatment rounds. The fertilizer treatment has no effect on adoption of Western Seed Company maize hybrids. The seed treatment has no effect on use of inorganic fertilizer use on maize. The fertilizer treatment increases fertilizer use on maize substantially; given that fertilizer use is measured in IHST of kilograms per acre, the point estimate of 0.99 implies roughly a doubling of fertilizer use on average across the post-treatment rounds.

In the transitional zone of Western Kenya, the seed treatment increases adoption of Western Seed Company maize hybrids by 9 percentage points on average across the post-treatment rounds. This relatively low net compliance is due in part to the high rate of adoption of Western Seed Company maize hybrids in the control group. The seed treatment does not have a statistically significant effect on use of inorganic fertilizer use on maize. The fertilizer treatment has a positive but not statistically significant effect on use of inorganic fertilizer use on maize. The smaller magnitude and statistical significance of this effect in the transitional zone relative to the mid-altitude zone is due in part to the relatively high use of fertilizer in the transitional zone absent treatment.

In the transitional zone of Central Kenya, the seed treatment increases adoption of Western Seed Company maize hybrids by 7 percentage points on average across the post-treatment rounds. The seed treatment does not have an effect on fertilizer use that differs from zero with statistical significance, which is not surprising given that fertilizer use is high in the transitional zone absent treatment.

Table A2: Effects on input use.

	Mid-Altitude		Transitional-Western		Transitional-Central	
	WSC	*Fertilizer*	*WSC*	*Fertilizer*	*WSC*	*Fertilizer*
Seed Treatment	0.16***	-0.02	0.09	0.10	0.07***	-0.18
	(0.03)	(0.27)	(0.07)	(0.22)	(0.02)	(0.17)
Fertilizer Treatment	-0.00	0.99***	-0.06	0.31		
	(0.01)	(0.15)	(0.04)	(0.23)		
Seed*Fertilizer Treatment	-0.00	-0.12	0.03	0.00		
	(0.04)	(0.28)	(0.06)	(0.28)		
Control mean	0.02	1.63	0.19	3.73	0.01	4.12
Observations	1178	1178	856	856	1016	1016
R-squared	0.11	0.34	0.06	0.06	0.05	0.07
F-statistic	5.82	90.40	5.98	8.18	2.81	2.92

WSC outcome variable is an indicator of adoption of Western Seed maize hybrids (0/1).
Fertilizer outcome variable is inorganic fertilizer used on maize (IHST of kg/ac).
Observations are of households observed in all rounds and pool both rounds post-treatment.
Pair indicator variables included as controls. Standard errors clustered by village.
* = 10% significance, ** = 5%, *** = 1%

Appendix E: Balance Checks in Western and Central

Table A3 estimates how baseline characteristics differ by treatment status in Western. Table A4 estimates how baseline characteristics differ by treatment status in Central.

Table A3: Balance at baseline, Western (N=1017).

	Summary Stats		Estimates from OLS		
	Pooled	*Control*	*Seed*	*Fert*	*Seed*Fert*
Hybrid main seasons (0-1)	0.51	0.52	0.00	-0.06**	0.06
	(0.44)	(0.45)	(0.04)	(0.03)	(0.04)
Fertilizer main seasons (0-1)	0.49	0.53	-0.02	-0.07**	0.03
	(0.46)	(0.46)	(0.04)	(0.03)	(0.05)
Dry maize yield (kg/ac)	368.49	375.01	7.50	-12.75	-18.17
	(425.64)	(428.83)	(37.21)	(26.93)	(40.24)
Acres (maize)	1.32	1.38	-0.26***	0.03	0.19
	(1.17)	(1.32)	(0.09)	(0.10)	(0.12)
Acres (total)	1.76	1.82	-0.27**	0.06	0.16
	(1.51)	(1.59)	(0.12)	(0.12)	(0.15)
Income per capita (100 ksh)	279.60	252.91	25.31	10.52	38.22
	(391.37)	(347.63)	(34.66)	(30.42)	(43.33)
Poverty (0-1)	0.32	0.33	-0.02	-0.00	0.01
	(0.23)	(0.24)	(0.02)	(0.02)	(0.03)
Dietary diversity (0-12)	6.68	6.52	0.28**	0.07	-0.03
	(1.62)	(1.65)	(0.14)	(0.14)	(0.19)
Food insecure (0/1)	0.65	0.70	-0.07**	-0.03	0.00
	(0.48)	(0.46)	(0.04)	(0.04)	(0.06)

Pooled and Control report means (standard deviations). Seed, Fert, and Seed*Fert report point estimates obtained by OLS with pair indicators as controls (standard errors clustered by village). Significance: * = 10%, ** = 5%, *** = 1%

Table A4: Balance at baseline, Central (N=508).

	Summary Stats		**OLS**
	Pooled	*Control*	*Seed*
Hybrid main seasons (0-1)	0.83	0.84	-0.03
	(0.34)	(0.32)	(0.04)
Fertilizer main seasons (0-1)	0.92	0.92	-0.01
	(0.29)	(0.30)	(0.03)
Dry maize yield (kg/ac)	428.46	432.64	-5.14
	(489.72)	(515.44)	(39.63)
Acres (maize)	0.76	0.74	0.04
	(0.70)	(0.70)	(0.08)
Acres (total)	1.29	1.27	0.04
	(1.14)	(1.08)	(0.14)
Income per capita (100 ksh)	513.65	430.68	167.41**
	(740.83)	(554.21)	(70.35)
Poverty (0-1)	0.13	0.13	-0.01
	(0.17)	(0.18)	(0.01)
Dietary diversity (0-12)	7.63	7.63	-0.01
	(1.45)	(1.40)	(0.15)
Food insecure (0/1)	0.44	0.46	-0.03
	(0.50)	(0.50)	(0.05)

Pooled and Control report means (standard deviations). Seed reports point estimates obtained by OLS with pair indicators as controls (standard errors clustered by village).
Significance: * = 10%, ** = 5%, *** = 1%

Appendix F: Randomization Inference Tests

In order to conduct randomization inference tests, we define a set of three mutually exclusive dummy variables: D^S indicates those farms that were assigned to the seed treatment but lost the fertilizer lottery; D^F are farms that won the fertilizer lottery but were not assigned to the seed treatment; and, D^{SF} are farms that were assigned to the seed treatment and won the fertilizer lottery. The specification for farm i in village v, site s, matched pair p and time period t is:

$$y_{ivspt} = \beta_0 + \beta_1 y^0_{ivsp} + \alpha^S D^S_{is} + \alpha^F D^F_{is} + \alpha^{SF} D^{SF}_{is} + \left[\mu_p + \varepsilon_{ivspt}\right], t = 1, 2 \tag{5}$$

where y^0_{ivsp} is baseline maize yields and μ_p is a matched pair fixed effect included to account for stratification of the seed treatment following Bruhn and McKenzie (2009). The error term (ε_{ivspt}) is clustered by village, the level of stratification for the seed treatment. In the central region, where there was no fertilizer lottery, we estimate Eq. (5) after eliminating the terms involving D^F and D^{SF}.

In the western region, the randomization inference procedure re-randomizes the seed and fertilizer treatments 1000 times and determines the share of these replications in which the estimated effects of the re-randomized treatments are greater in absolute value than the estimates that we obtain from our sample. Each re-randomization follows the stratification strategy of our experimental design, with seed randomized within matched pairs and fertilizer randomized within villages.

In the central region, the experiment only involves the seed treatment. This allows us to replicate all 64 possible combinations of seed treatment assignments stratified by matched pair. The randomization p-value is then the share of these replications that produce an estimate as large as the estimate from our sample in absolute value.

Table A5 gives the estimates of Eq. (5) with p-values based on clustered standard errors in parentheses and p-values based on randomization inference in brackets. These results our analogous to the main results in our paper in Table 3; parameter estimates and p-values based on clustered standard errors for “Seed Treatment Only” and “Fertilizer Winner Only” are identical to those for “Seed Treatment” and “Fertilizer Treatment” in Table 3. Randomization inference supports our conclusion that Western Seed varieties have a meaningful impact in the mid-altitude zone, but not in the transitional zones. For the mid-altitude zone, p-values from randomization inference indicate that our estimated treatment effects of the seed and fertilizer have strong statistical significance, with only 1% of replications obtaining larger estimates than the estimates that we obtain from our sample.

Table A6 gives the estimates of Eq. (5) for the heterogeneity analysis in the mid-altitude zone conducted in Section 5. These results our analogous to Table 4; parameter estimates and p-values based on clustered standard errors for "Seed Treatment Only" and "Fertilizer Winner Only" are identical to those for "Seed Treatment" and "Fertilizer Treatment" in Table 4. For hybrid users, p-values from randomization inference indicate strong statistical significance of the seed and fertilizer treatment comparable to the significance suggested by our clustered standard errors. For hybrid non-users, p-values from randomization inference indicate statistical significance of the seed and fertilizer treatments at the 5% significance level, a much stronger result than when using p-values based on clustered standard errors. This suggests that, for hybrid non-users, our treatment effect estimates are large relative to the uncertainty from treatment assignment within the sample, but are not large relative to the uncertainty from sampling from the general population.

Table A5: Effects on maize yield (IHST of kg/ac).

	Mid-Altitude	**Transitional**	
	All Farms	*Western*	*Central*
Seed Treatment Only, $\hat{\alpha}^S$	0.23	-0.17	0.10
	(0.08)	(0.15)	(0.50)
	[0.01]	[0.07]	[0.69]
Fertilizer Winner Only, $\hat{\alpha}^F$	0.23	0.11	
	(0.04)	(0.25)	
	[0.01]	[0.25]	
Seed Treatment & Fertilizer Winner, $\hat{\alpha}^{SF}$	0.23	0.13	
	(0.05)	(0.16)	
	[0.01]	[0.15]	
Baseline Yield, $\hat{\beta}_1$	0.11	0.15	0.20
	(0.00)	(0.00)	(0.00)
Observations	1178	856	1016
R-squared	0.07	0.06	0.11

Results obtained by ordinary least squares estimation.

All specifications include pair indicator variables as controls.

P-values from clustering standard errors at the village-level in parentheses.

P-values from randomization inference in brackets.

Table A6: Heterogeneous impact of locally adapted improved varieties in the mid-altitude zone (IHST of kg/ac).

	Hybrid Non-Users		**Hybrid Users**	
	Pooled	*Endline*	*Pooled*	*Endline*
Seed Treatment Only, $\hat{\alpha}^S$	0.20	0.26	0.40	0.78
	(0.19)	(0.08)	(0.02)	(0.00)
	[0.04]	[0.06]	[0.02]	[0.01]
Fertilizer Winner Only, $\hat{\alpha}^F$	0.22	0.18	0.35	0.54
	(0.11)	(0.21)	(0.04)	(0.02)
	[0.03]	[0.19]	[0.06]	[0.03]
Seed Treatment & Fertilizer Winner, $\hat{\alpha}^{SF}$	0.21	0.23	0.49	0.88
	(0.09)	(0.05)	(0.02)	(0.00)
	[0.04]	[0.10]	[0.00]	[0.00]
Baseline Yield, $\hat{\beta}_1$	0.12	0.09	0.02	-0.02
	(0.00)	(0.03)	(0.69)	(0.72)
Observations	964	482	208	104
R-squared	0.06	0.03	0.19	0.22

Results obtained by ordinary least squares estimation.

All specifications include pair indicator variables as controls.

P-values from clustering standard errors at the village-level in parentheses.

P-values from randomization inference in brackets.

Appendix G: Impacts on Net Revenue Per Acre by Agro-ecological Zone

For completeness, this appendix reports estimates of the financial impacts of the interventions when we explore impacts by agro-ecological zone, akin to Table A5 in the main body of the paper. Consistent with these zone level yield results, Table A7 shows a positive (but statistically insignificant) gain from the treatment in the mid-altitude zone and no economic gain from the treatment in the transitional zones.

Table A7: Impacts on net revenue per acre by agro-ecological zone (IHST of ksh/ac).

	Mid-Altitude	**Transitional**	
	All Farms	*Western*	*Central*
Seed Treatment, $\hat{\boldsymbol{\delta}}_{\mathbf{1}}$	0.26	-0.63**	-0.04
	(0.22)	(0.29)	(0.33)
Fertilizer Treatment, $\hat{\boldsymbol{\delta}}_{\mathbf{2}}$	0.11	0.05	–
	(0.23)	(0.18)	–
Seed*Fertilizer Treatment, $\hat{\boldsymbol{\delta}}_{\mathbf{3}}$	-0.26	0.75**	–
	(0.30)	(0.30)	–
Baseline Economic Yield, $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$	0.05	0.06	0.14***
	(0.03)	(0.04)	(0.04)
Percent effects			
Seed Treatment	27%	-49%	-9%
Fertilizer Treatment	9%	4%	–
Control mean	8.94	9.91	8.60
Observations	1178	856	1016
R-squared	0.02	0.03	0.06

Pair indicator variables included as controls. Standard errors clustered by village.
* = 10% significance, ** = 5%, *** = 1%

Appendix H: The Fertilizer-Seed Interaction Puzzle

As discussed in Sections 4 and 5 in the main body of the paper, households that received both the seed treatment and won the midline fertilizer lottery, seem not to have gained from the latter additional intervention. In contrast, households in control areas that did not receive the seed treatment but who did win the fertilizer lottery appear to have benefited substantially from the grant of fertilizer. Table A8 shows that at least part of this puzzling finding is apparently due to a greater sharing of fertilizer by grant recipients in seed treatment areas. As can be seen, the 50 kg grant of fertilizer to lottery winners in control areas (no seed treatment), increased their mean midline fertilizer use from 9 kg to 44 kg. In seed treatment areas, midline fertilizer use by lottery winners increased by 10 kg less (increasing to only 34 kg).[29] This finding suggests that fertilizer lottery winners in seed treatment areas shared more of the windfall fertilizer grant with their neighbors. This suggestion is corroborated by the figures in Table A8 on midline fertilizer use by households that did not win the fertilizer lottery.[30]

We see a similar pattern if we simply look at fertilizer used on maize (Table A9). The

---

[29] The figures reported in Table A8 are means from the different survey rounds. A regression analysis of the impact of the treatments on fertilizer use that mimics Eq. (5) gives estimates that are nearly indistinguishable from the differences shown in the table.

[30] Note that all fertilizer lottery winners were survey respondents, whereas communities contained many non-surveyed farmers who did not win the fertilizer lottery. It is thus not surprising that the measured spillover increments to lottery losers is smaller than the apparent net sharing by lottery winners.

Table A8: Fertilizer use (kg) by survey round and treatment assignment in the mid-altitude zone (sub-sample means).

Survey Round	**No Seed Treatment**		**Seed Treatment**	
	Lottery Losers	*Fertilizer Lottery Winners*	*Lottery Losers*	*Fertilizer Lottery Winners*
Baseline	9.4	9.0	12.4	8.1
Midline	19.2	44.3	24.7	34.3
Endline	14.1	18.7	16.9	15.3

Table A9: Fertilizer use on maize (kg) by survey round and treatment assignment in the mid-altitude zone (sub-sample means).

Survey Round	**No Seed Treatment**		**Seed Treatment**	
	Lottery Losers	*Fertilizer Lottery Winners*	*Lottery Losers*	*Fertilizer Lottery Winners*
Baseline	8.1	7.3	11.4	7.4
Midline	17.2	40.0	23.2	31.9
Endline	13.2	16.1	15.4	13.7

gap between fertilizer lottery winners and losers in seed treatment communities is about 1 kg smaller (8.7 versus 9.6) when looking at fertilizer used on maize versus fertilizer used on all crops. It is of course this differential use of fertilizer driven by the lottery that is the basis for identifying the impact of winning the lottery in seed treatment communities.

One explanation for this greater sharing of windfall fertilizer in seed treatment areas is that the seed treatment information campaign in treatment sites emphasized the importance of fertilizer to improving maize yields (especially when applied to hybrids). It would therefore not be surprising if fertilizer lottery losers in seed treatment sites put greater pressure on winners to share their windfall fertilizer than losers would have done in control sites. Given this sharing pattern, it is unsurprising that the added benefit of winning the fertilizer lottery in seed treatment areas was smaller than the benefit in seed control areas. Specifically, note that even if fertilizer lottery winners in seed treatment communities gave away 10 kg more of their windfall fertilizers, the data still indicate that these winners used 9 kg more fertilizer on maize than their lottery-losing neighbors (32 kg versus 23 kg). In control communities (no seed treatment) the fertilizer use gap on maize between lottery winners and losers was two and a half times higher than in seed treatment communities (winners used 40 kg versus losers who used 17 kg).

So how much of the seed-fertilizer interaction puzzle can in principal be explained by this pattern of differential sharing? Using the stylized parameters in Table A1, we can roughly gauge the likely impact of this differential sharing on the yield differential between fertilizer lottery winners and losers in both seed treatment and control areas. Those parameter values assume that 1 kg of fertilizer returns 5 kg additional maize yield when fertilizer is applied to local varieties, whereas it returns 25 kg of additional maize when applied to locally adapted improved varieties. Using these and the other values in the table, we find that winning the fertilizer lottery in seed control areas would have boosted yields by 37% given the observed patterns of fertilizer use and sharing (note that Tables 3 and 4 estimate percentage changes). In contrast, the midline yield boost for fertilizer lottery winners in seed treatment areas would have been only 19% after taking into account the partial compliance in seed treatment areas.[31] If we perform the same calculations for the endline year (using the figures for endline fertilizer use in Table A8 which show that differences between lottery winners and losers largely disappear) we find that the differences are smaller than at midline (8% for the seed control areas and -2% for the seed treatment areas).

The estimates in Table 3 based on the actual experiment pool data across the midline and endline seasons and hence estimate an average impact of the different treatments across these two seasons. Averaging our back of the envelope calculations for the impact of winning the fertilizer lottery across the two years implies a 23% gain in seed control areas and 8% in seed treatment areas. The 23% calculated gain in control areas is an artifact of our assumption about the fertilizer responsiveness of traditional varieties, but almost exactly matches the 25% statistical estimate shown in Table 3.[32] The 8% gain for lottery winners in seed treatment areas is larger than the 0% estimate implied the estimated coefficients in Table 3, although the 95% confidence interval around the 0% point estimate would clearly contain 8% given the magnitude of the estimated standard errors.

While this pattern of differential sharing seems able to explain much of the fertilizer-seed interaction puzzle, there are other possible, non-mutually exclusive explanations. One is that the shared fertilizer went to those with the most experience and greatest capacity to properly apply and benefit from fertilizers (who could be imagined to most heavily demand the windfall fertilizer from lottery-winning neighbors). In contrast, lottery winners were randomly selected farmers, many of whom had never used fertilizers on their farms. This pattern could further explain the reason why fertilizer lottery winners did not seem to benefit as much as would be expected. Unfortunately, our data do not permit any further analysis of this question.

---

[31] In conformity with the figures in Table 2, we assume that only 20% of farmers in seed treatment areas adopt the locally adapted hybrids, with the remainder using retained local varieties. The partial adoption of improved varieties would lower this differential further.

[32] For example, if we assume that the fertilizer responsive of local seeds is 10 kg of maize per-kg of fertilizer, the calculated yield difference between lottery winners and losers increases to 33% rather than 23%.