Relaxing multiple agricultural productivity constraints at scale

Joshua W. Deutschmann
Maya Duru
Kim Siegal
Emilia Tjernström

August 19, 2024

Abstract

No single constraint can explain the stagnant agricultural productivity growth in sub-Saharan Africa. Most interventions that relax individual barriers to productivity have delivered disappointing results. We evaluate an at-scale program that targets several productivity constraints with a bundled intervention, using a randomized controlled trial in western Kenya. Program participation increases maize yields by 26%, total maize output by 24%, and profits by 18%. While we cannot directly test whether the program's success is due to its bundled nature, we find patterns in the data that are consistent with this hypothesis.

Deutschmann: University of Chicago, jdeutschmann@uchicago.edu. Duru: Office of Evaluation Sciences, U.S. General Services Administration, maya.joan.duru@gmail.com. Siegal: Mathematica, ksiegal@mathematica-mpr.com. Tjernström: Monash University, emilia.tjernstrom@monash.edu. Duru and Siegal were employed by One Acre Fund at the time of this study. This paper previously circulated under the title "*Can smallholder extension transform African agriculture?*" The authors thank Kibrom Abay, Brad Barham, Chris Barrett, Lori Beaman, Leah Bevis, Michael Carter, Lorenzo Casaburi, Travis Lybbert, Nick Magnan, Jeremy Magruder, William Masters, Laura Schechter, Andrew Simons, Jeffrey Smith, Tavneet Suri, and Chris Udry for helpful comments and suggestions. We have also benefited from comments by seminar and conference participants at CSAE, the IDEAS Summer School in Development, the NBER & African Development Bank Transforming Rural Africa Conference, NEUDC, Y-RISE, and UW-Madison. All remaining errors are our own. The views expressed in this paper are those of the authors and do not reflect the official policy or position of the U.S. Government. This study received financial support from the Global Innovation Fund. Strathmore University's IRB approved the study (IRB Approval Number: SU/IRB 0062/16). AEA Social Science Registration: AEARCTR-0006675.

1 Introduction

How should scarce public resources be allocated across sectors of the economy to bolster economic growth and development? This question occupies a central role in development economics and policy debates. As a large contributor to GDP, employment, and food security, the agricultural sector is often at the heart of such discussions. A growing body of evidence demonstrates that agricultural productivity plays a key role in the economic development process by reducing poverty (see de Janvry and Sadoulet 2010 for a review), increasing national welfare (Bravo-Ortega and Lederman, 2005; Gollin, Hansen, and Wingender, 2021; Ligon and Sadoulet, 2011; Ravallion and Chen, 2007), and driving structural change (Jayne, Chamberlin, and Benfica, 2018a; McArthur and McCord, 2017).

Despite non-trivial government investments in the agricultural sector, agricultural productivity in sub-Saharan Africa (SSA) lags other regions of the world, both in terms of levels and in growth (Block, 2014; Wollburg, Bentze, Lu, Udry, and Gollin, 2024; World Bank, 2008).[1] This has led some to question the viability of agriculture-led growth strategies in the region. Further, opinions differ on whether African governments should continue to direct resources towards the smallholders that farm the majority of the region's arable land, or if medium- and large-scale farms would be more cost-effective.[2] In recent years, many African governments have shifted resources towards developing medium- and large-scale farms (Jayne, Chamberlin, Traub, Sitko, Muyanga, Yeboah, Anseeuw, Chapoto, Wineman, Nkonde, and Kachule, 2016).

We contribute to this policy discussion by answering a set of slightly narrower research questions, which we consider fundamental for answering these bigger-picture policy questions. First, we ask whether a scaled-up agriculture-focused program for African smallholders can generate economically meaningful productivity gains. Having answered this question in the affirmative—a pre-condition for the viability of smallholder agriculture-led growth strategies—we try to understand whether specific features of the program that we study can shed light on the relative importance of different market failures in the context of our study. Further, we examine treatment effect heterogeneity to understand whether the program is more effective for certain types of farmers, and move beyond yields to look at whether the net effects on participant welfare are positive. Finally, we try to answer

[1] Suri and Udry (2022) calculate that around 5 percent of all government expenditures across Africa go to agriculture, including investments in extension services, input subsidies, and rural road infrastructure.

[2] Smallholder agriculture constitutes a large fraction of the region's total agricultural production and its farmers make up a substantial portion of the population. This pattern holds true for many low-income countries: Lowder, Skoet, and Raney (2016) estimate that 72% of farms in the world use less than one hectare of land.

the broader question of whether the program increases social welfare, by comparing our estimated benefits to the program's costs.

Our primary contribution is to provide experimental evidence of the productivity impacts of an at-scale program run by One Acre Fund (1AF), an established non-governmental organization (NGO). 1AF was founded in 2006 and currently works with over one million farm households in seven countries across Eastern and Southern Africa. We evaluate their "full service" smallholder farmer program, which offers enrolled farmers a bundle that includes loans for improved seeds and fertilizer, training on modern agricultural techniques, and input insurance. We find that 1AF program participation causes statistically and economically significant increases in productivity and total output, defined as per-acre maize yields and total maize output at the farm level, respectively. Specifically, the program increases maize yields by 26 percent and maize output by 24 percent. Our identification strategy relies on a cluster-randomized experiment, in which we randomized access to the program for farmer groups in western Kenya.

Moving beyond agricultural productivity, we seek to understand whether the program increases farmer welfare. Using rich plot-level data on input use and farming practices, we find that program participation changes several different dimensions of farming behavior. We assign a monetary value to the main costly inputs, including labor, fertilizer, and purchased seed, and subtract them from the market value of the maize outputs to obtain a measure of farmer profits. On average, we find that program participation is profitable, with the net value of maize production increasing by about 18%. Although we cannot fully capture welfare changes from other income sources, the median farmer in our sample dedicated all of their cultivated acreage to maize at baseline. This suggests that changes in the net value of maize production constitute an important part of agricultural income, and likely overall household income.

We then explore whether the average treatment effects mask underlying impact heterogeneity, using a variety of heterogeneity analyses. Treatment effect heterogeneity could indicate potential efficiency gains from programmatic changes or improved participant targeting. We find little evidence of systematic heterogeneity in treatment effects on productivity or output. There are several potential explanations for this lack of heterogeneity, including the potential screening effect of the small upfront fee that 1AF requires prior to enrollment—an important difference between 1AF and most other agricultural productivity programs.[3]

[3] Studies of health products in low-income countries generally find cost-sharing mechanisms to be ineffective screening devices (see for example Ashraf, Berry, and Shapiro 2010; Cohen and Dupas 2010; Tarozzi, Mahajan, Blackburn, Kopf, Krishnan, and Yoong 2014). However, Beaman, Karlan, Thuysbaert, and Udry

Another contribution of our study is that we take seriously the broader question of whether the benefits that we describe above are worth the overall costs of the 1AF intervention. Specifically, we conduct a cost-benefit analysis to determine the net economic impact of the 1AF program in western Kenya. Rather than defend a particular set of assumptions, we account for different types of uncertainty using Monte Carlo simulations. The simulations produce distributions of the net benefits that emerge under a range of different assumptions. This analysis suggests that program participation for farmers who are new to the program generates an average net social benefit of $36.6, with a standard deviation of 33.7. For farmers who have prior experience with the program (pre-exposed farmers), the simulated distribution shows an average net social benefit of $18, with a standard deviation of 26.4. Further, we obtain positive net social benefits for 86 and 75 percent of cases, respectively, for new and pre-exposed farmers.

This paper contributes to the broad literature that aims to quantify the productivity impacts of programs that relax constraints to technology diffusion and productivity growth on smallholder farms.[4] This body of research includes studies of agricultural extension policies and input subsidy programs, with interventions ranging widely in scope and scale. A substantial proportion of the studies in this literature report disappointing results, whereby even meaningful improvements in farmer information access, practices, or input use fail to translate to measurable increases in yields (Cole and Fernando, 2021; Udry, di Battista, Fosu, Goldstein, Gurbuz, Karlan, and Kolavalli, 2019) or profits (Beaman, Karlan, Thuysbaert, and Udry, 2013).

A notable subset of this literature focuses on input subsidy programs (ISP), which have become a major component of agricultural policy for at least ten countries in sub-Saharan Africa in the last two decades (Jayne and Rashid, 2013). The evidence on the productivity impacts and cost-effectiveness of these programs is mixed. In a review of nearly 80 studies of ISPs, Jayne, Mason, Burke, and Ariga (2018b) conclude that their overall productivity impacts tend to be underwhelming, in part due to low observed crop-yield returns to fertilizer. Tamim, Harou, Burke, Lobell, Madajewicz, Magomba, Michelson, Palm, and Xue (2024) revisit the study population of a field experiment in Tanzania, which provided a yield- and profit-increasing fertilizer subsidy. Five years on, the program impacts had dissipated such that treatment farmers were using no more fertilizer than they did prior to the intervention, and no more than their control group counterparts. In contrast, Carter, Laajaj, and Yang (2021) estimate that a one-off input subsidy program in Mozambique increased maize yields by 23 percent, that program benefits well outweighed costs, and that

(2023) find evidence of positive selection into an agricultural credit program in Mali.

[4]See Magruder (2018) and Suri and Udry (2022) for recent reviews.

impacts persisted the year after the subsidy ends.[5] Similarly, Fishman, Smith, Bobić, and Sulaiman (2022) find persistent effects of an intervention in Uganda that provided farmers with samples of improved seeds and information on improved agricultural practices after the program was phased out.

The literature commonly highlights credit, risk, and information as key barriers to adoption of improved agricultural technologies (Feder, Just, and Zilberman, 1985; Magruder, 2018). The varying results across studies, including ours, may reflect differences in which constraints or market failures are most relevant in each context. For example, we might expect input subsidy programs to have persistent effects if they primarily relax an informational constraint to adoption, enabling farmers to learn about the profitability of improved agricultural inputs. However, if credit or risk constraints bind, then temporary subsidies or other information interventions may not be adequate policy instruments, as in Tamim et al. (2024) and Cole and Fernando (2021).

The 1AF program targets several of the above constraints at once: input loans for high-quality seeds and fertilizer serve to relax liquidity constraints, input insurance aims to reduce the risk of making profitable investments, and training on improved farming practices relaxes information constraints. None of the program ingredients are novel on their own; in fact, they overlap substantially with the kinds of programs that the World Bank promoted as part of their agricultural extension investments in the 1960s and 1970s (Birkhaeuser, Evenson, and Feder, 1991; Chambers, 1983). However, the program's tight bundling of its component parts into a single unified package (Tinsley and Agapitova, 2018) distinguish it from most other programs. This bundled approach is not typical of government-led extension, nor is it common in the economics literature, which has tended to evaluate the impact of relaxing a single constraint at a time.[6]

Our results are consistent with the notion that multiple constraints matter for farmers' decisions and outcomes in our study context, which in turn suggests that the bundled nature of the 1AF program might play an important role in its success.[7] We acknowledge that bundling complicates inference on the underlying economic mechanisms and does not allow

[5] Diop (2024) uses variation in the roll-out of Zambia's ISP to estimate its returns and finds that the large-scale program increases yields by around 15 percent. However, the study does not examine farmer profits or program costs, which makes it hard to evaluate whether the ISP is cost-effective.

[6] Some recent papers show promising impacts of interventions that target multiple constraints in the context of commercially-focused smallholder agriculture (Arouna, Michler, and Lokossou, 2021; Deutschmann, Bernard, and Yameogo, 2023; Macchiavello and Miquel-Florensa, 2019; Park, Yuan, and Zhang, 2023).

[7] Note that the Kenyan context differs rather dramatically from the settings in Carter et al. (2021) and Fishman et al. (2022), where study participants have minimal experience with improved agricultural inputs. Nearly all of our study participants have prior experience using hybrid seeds and fertilizer on their fields.

us to cleanly separate out the impact of different components of the bundle. Nonetheless, our data allow us to examine several farming choices that precede productivity and output along the causal chain. The resulting empirical patterns suggest that both information and credit constraints are binding in the study context. Our results echo recent findings from multi-faceted anti-poverty programs, where individual intervention components seem unable to generate the same magnitude of effects.[8]

An obvious limitation of our study is that it focuses on a single crop in one agricultural region in one season, suggesting some caution in generalizing our results to other seasons and other populations (Rosenzweig and Udry, 2020). However, our results are consistent with 1AF's non-experimental monitoring and evaluation results, and consistent with farmers' own revealed preference, with the program continuing to grow rapidly and many farmers choosing to re-enroll (Deutschmann and Tjernström, 2018; One Acre Fund, 2020). Taken together, our findings on the broad impacts of the program and its cost-effectiveness suggest there is indeed a role for targeting smallholder farmers in an agriculture-led growth strategies. Supporting and learning from effective programs operating at scale will be a key component of any such strategy.

2 Context, data, and experimental design

2.1 One Acre Fund's program in Kenya

We analyze the main operating model of an established agricultural NGO, One Acre Fund (1AF). Founded in 2006, the organization has grown rapidly in the last several years: enrollment has grown from 200,000 farm households in 2014 to more than one million in 2020 (One Acre Fund, 2020). 1AF's "full service program" provides farmer groups with a package of support that includes loans for improved seeds and fertilizer, convenient distribution of those seeds and fertilizer, regular training on modern agricultural techniques, input and funeral insurance, and marketing advice to encourage farmers to obtain higher prices for their output (Tinsley and Agapitova, 2018). In our study context, 1AF provides input credit as joint liability loans to farmer groups, organized by geographical area and typically comprised of 8-12 farmers.

The program's focus on credit, risk, and information has support in the economics

[8] The theoretical literature on poverty traps is a notable exception to the frequent focus on single constraints. A growing empirical body of evidence complements this theoretical literature by investigating the possibility that poor households may need large, bundled interventions in order to move out of poverty (Balboni, Bandiera, Burgess, Ghatak, and Heil, 2022; Bandiera, Burgess, Das, Gulesci, Rasul, and Sulaiman, 2017; Banerjee, Duflo, Goldberg, Karlan, Osei, Pariente, Shapiro, Thuysbaert, and Udry, 2015).

literature, which has accumulated substantial evidence that failures in these domains hinder farmers' ability or willingness to adopt improved agricultural technologies (Feder et al., 1985; Magruder, 2018). 1AF's decision to focus on providing high-quality inputs also has empirical support: a growing literature establishes that fertilizer and seeds sold in the region's local markets often fall short of quality standards, and that farmer perceptions of input quality matter even when inputs meet quality standards (Bold, Kaizzi, Svensson, and Yanagizawa-Drott, 2017; Hsu and Wambugu, 2022; Michelson, Fairbairn, Ellison, Maertens, and Manyong, 2021; Tjernström, Carter, and Lybbert, 2018).

Prior to enrolling in the smallholder farmer program, participants sort into self-selected farmer groups. Participants then choose how much of their land to enroll, in increments of 0.25 acres. The agricultural credit and input quantity provided by 1AF is determined as a function of land enrollment. Farmer groups are jointly responsible for repayment of these loans, which have relatively flexible loan terms, allowing repayment at any time during the growing season.[9] Groups must complete repayment in full within a two-week grace period after harvest, and most do: repayment rates have historically exceeded 97% (Tinsley and Agapitova, 2018).

To alleviate information constraints, 1AF field officers conduct repeated trainings and provide handouts on the benefits and proper use of improved inputs. To alleviate risk, 1AF provides yield-index insurance based on crop cuts and forgives a portion of the input loan in case of crop failure (Tinsley and Agapitova, 2018).

Our experiment took place in western Kenya, where 1AF has operated for more than ten years. Agriculture contributes 51% to Kenya's GDP and is dominated by small-scale producers (Government of Kenya, 2010). Despite the importance of the agricultural sector to the economy, most small-scale farmers are not running successful micro-enterprises. Households in Kenya typically derive their income from the production of a variety of crops (Sheahan, Black, and Jayne, 2013) and average productivity often fails to meet households' dietary needs (Kirimi, Sitko, Jayne, Karin, Muyanga, Sheahan, Flock, and Bor, 2011).

Although small-scale farmers in Kenya produce a range of crops across a diverse production environment, 1AF's program focuses on the dominant staple crop: maize. Maize is important both to the economy and for food security. Seventy percent of Kenya's maize is produced by smallholders who farm between 0.2 and 3 hectares (Government of Kenya, 2010, pg. 11-12). While Kenyan farmers use improved seed varieties and inorganic fertilizer at higher rates than neighboring countries, yields remain low. Further, the country remains a net importer of maize, despite policy targets to the contrary.

[9] Field, Pande, Papp, and Rigol (2013) suggest flexibility in loan repayment is important in encouraging illiquid investments among microentrepreneurs.

2.2 Experimental design

Recruitment, enrollment, and program implementation for this experiment took place in the Teso region of Kenya, following 1AF's standard protocol. Farmers who satisfied the basic program criteria paid a participation deposit of approximately $5 USD. To participate in the program, farmers must have a phone number and national identification. For the study, farmers had to meet two additional criteria: they had to consent to be part of the study and plan to cultivate at least 0.25 acres of maize. After contract signing, 1AF informed farmers about the study, that their participation would be voluntary, and went over informed consent.

Once participants had self-selected into groups of 8-12 farmers, we randomized treatment at the level of clusters, which consisted of 2-4 joint-liability farmer groups. This aggregation was designed to minimize potential spillovers by maximizing the distance between treatment clusters.

After participants had paid the fee and signed the contract, they were informed of the randomization, which took place as a public lottery. At this time, farmer groups who had been randomly assigned to the control group were informed that they would have to wait one year to enroll in the program. NGO staff offered them a roughly $10 USD discount for enrollment in the following season and a compensation package consisting of a bundle of household goods valued at the cost of participation.[10]

Dealing with sample contamination

While the ideal location for this study would arguably be an area where 1AF had never operated, this proved difficult in Kenya. Given 1AF's long presence in the country, all "untouched" regions would have been unrepresentative of the production environments and farmer populations that 1AF usually engages with. As a second-best option, the evaluation team decided to sample villages within district where 1AF was working, but where the program had not yet recruited farmers directly from those villages.

In practice, although 1AF had never offered its program to the sampled villages, about 64% of the farmers in our sample had managed to access it anyway, by "commuting" to neighboring villages to participate. The proportion of such "pre-exposed" farmers does not differ across treatment and control groups. We may nevertheless be concerned that the pre-exposed farmers introduce bias into our results. Accordingly, we report our main results both for the full sample and for a "primary sample," which refers to the smaller

[10]Specifically, households received a bag and thermos with a total value of less than $15 USD. These items were chosen because they are common in the study area and unlikely to have significant resale value.

sample of farmers who had never participated in 1AF programming.

The sign of any potential bias from including the pre-exposed farmers depends on at least two factors: the persistence of treatment effects and the correlation between selection and program returns. Suppose that 1AF participation changes farmers' production processes for the better. If pre-exposed farmers continue to operate more efficiently in subsequent seasons, then treatment effects estimated using the full sample may be attenuated.

In the other direction, the most eager farmers may have higher returns, in which case we would expect the impacts on our primary sample of hold-out farmers to be a lower bound of the true impacts. If, instead, these hold-out farmers resemble the never-adopters in Suri (2011) and have surprisingly high returns—but perhaps face high costs of participation—then it is possible that focusing solely on the primary sample might overestimate impacts.

2.3 Data

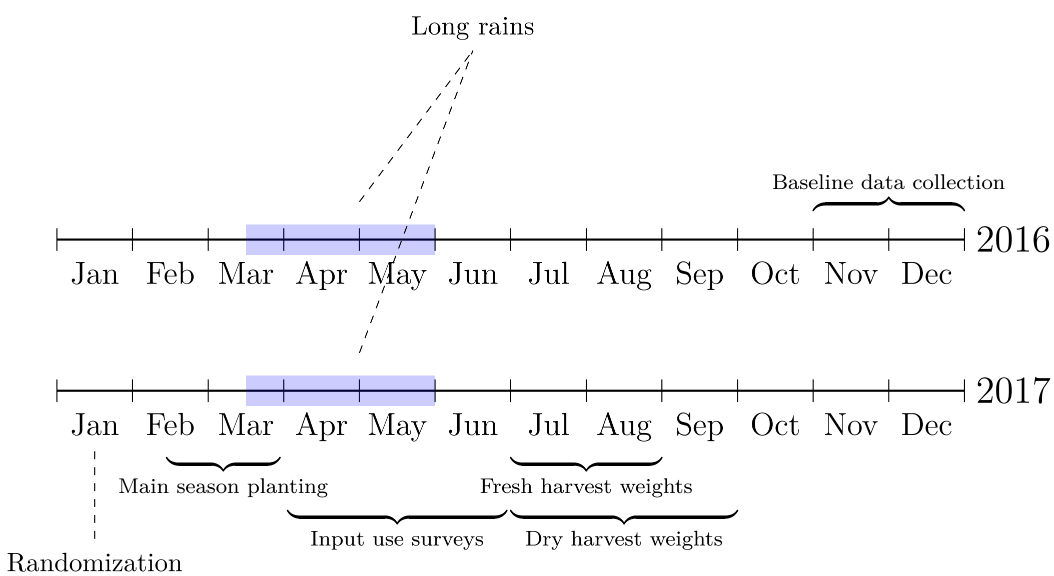


Figure 1: Timing of data collection activities and key features of the agricultural season

Baseline data collection took place in November and December of 2016—after initial enrollment but before contract signing. A public lottery assigned clusters of farmer groups to treatment in January 2017. Farmers in our sample typically planted their maize fields in February and March. Enumerator teams conducted input use surveys after the planting of the main season in 2017, from April through June. Enumerators collected fresh weight measurements between July-August 2017 and dry weight measurements between July-September 2017. Enumerators also conducted market surveys throughout the implementation of the project, collecting data on input prices from local agrovets, as well as maize and other output prices in local markets between May and October 2017.

2.3.1 Outcome variables

Our main analysis centers on maize yields and total maize output. Maize yields captures a direct comparison of average yields on the land treated farmers enroll in the 1AF program to the average yield in the control group. Since farmers rarely enroll all of their land, we might wonder about various forms of spillovers across a given farmer's plots. For example, there could be positive knowledge spillovers to non-enrolled plots, or participants may reallocate scarce complementary inputs such as labor to the enrolled plot. Conversely, farmers might prefer to spread 1AF-provided inputs across multiple plots in an effort to reduce risk. Therefore, we also measure and compare the maize yields that treated farmers obtain on their non-enrolled land to the yields on control farmers' maize plots. To further characterize the holistic impacts of program enrollment on farmers' agricultural production, net of substitution between plots, we generate and analyze an estimate of total maize output based on crop cut yields and total land cultivated.

To obtain crop-cut yields, enumerators collected and physically weighed fresh and dry harvests from two randomly placed 40-square-meter boxes. Following instructions in the survey instrument, enumerators placed the harvest box as follows: they started at a randomly-selected corner of the plot, took a randomly-selected number of steps perpendicular to the rows, followed by a random number of steps towards the center of the field. For treated farmers, enumerators also placed harvest boxes on unenrolled land if applicable. We measure the size of cultivated land using GPS readings, with enumerators walking the boundaries of each plot three times.[11] Enumerators measured treated farmers' enrolled and unenrolled land separately. Based on these objective measures of land size and crop cut yields, we generate a measure of total maize output. Appendix B provides additional variable construction details.

To further quantify the potential welfare effects of the program, we calculate profits as the value of output less farmers' costs (see Appendix B for more details on the construction of these variables). Revenues are the product of total output and average market prices from nearby vendors. We believe this to be a conservative estimate of revenues, as the 1AF

[11] A growing body of research documents non-classical measurement error in self-reported land size and harvests in developing-country studies (see e.g. Abay, Bevis, and Barrett 2021; Carletto, Savastano, and Zezza 2013; Desiere and Jolliffe 2018; Gourlay, Kilic, and Lobell 2019). Self-reported harvests are additionally subject to recall bias. The literature generally finds that crop-cut measurements are a more reliable way to measure physical yields than relying on self-reports.

program encourages farmers to store some maize for later sale when prices may be higher. The program additionally sells hermetic PICS bags to farmers, which can reduce post-harvest losses and quality degradation, as well as potentially allowing farmers to postpone sales to a time when prices are higher Burke, Bergquist, and Miguel (2019). Administrative data from 1AF suggests about 10% of farmers purchased at least one PICS bag. The returns to quality and impact of PICS bag adoption on storage are likely to be small in practice (Bold, Ghisolfi, Nsonzi, and Svensson, 2022; Channa, Ricker-Gilbert, Feleke, and Abdoulaye, 2022), but we conservatively assume no difference in price earned by treated and control group farmers.

To estimate costs, we combine input costs, labor costs, land rental costs, and program enrollment costs. For inputs, we elicit the quantity of inputs farmers applied to their plots. For farmers in the treatment group, we further distinguish between inputs applied to enrolled and non-enrolled land. To attribute costs to the inputs that farmers used, we rely on two sources of data. For inputs sourced from 1AF, we rely on 1AF administrative data that detail the prices farmers paid, which are standard across the sample. For all other inputs that treatment and control farmers used, we assign prices based on mean prices measured in a sample of local markets. Overall, a comparison of input prices across these two datasets confirms that 1AF charges seed and fertilizer prices that are comparable to local market prices. We additionally account for program enrollment costs, which include input delivery costs and insurance premiums.[12]

Labor costs are challenging to measure in this context, given the prevalence of unpaid labor. To reduce recall bias, we elicit early-season paid and family labor use in a survey administered shortly after planting (including self-reported labor for land preparation, plowing, and planting). Late-season paid and family labor use, including for weeding and harvest, was collected shortly after harvest. We ascribe a monetary value to family (unpaid) labor by calculating the mean day wage reported within the sample. We show two estimates of profit: one in which we value unpaid labor at 50% of the paid day wage (based on rural unemployment rates in Kenyan DHS data) and one in which we value unpaid labor at the full day wage.[13] We then multiply these labor prices by the number of person-days of unpaid labor.

[12] There were no insurance payouts reported in the year of our study among our sample. We conservatively do not ascribe an expected benefit to the input insurance.

[13] Our approach to valuing unpaid labor matches the approach in Agness, Baseler, Chassang, Dupas, and Snowberg (2022), which suggests valuing unpaid labor at 60% of the paid labor day wage.

2.3.2 Attrition

Attrition in our data is mostly due to missing variables, not to participants dropping out of the sample entirely. During data cleaning, we are primarily forced to drop observations for missing land size and harvest data. Land size data was collected during and immediately after harvest time, and in some cases enumerators could not complete the land survey on some or all plots. This issue affects 142 farmers in our sample, and occurs slightly more often among treatment farmers (7%) than among control group farmers (4%).

The increased attrition among treatment group farmers is primarily driven by missing land measurement on non-enrolled land. Enumerators recorded harvest data both immediately after harvest (the "fresh weight") and after drying (the "dry weight"). Our analysis relies on dry weight estimates of yields, which are a more comparable measure across farmers since moisture content before drying can vary. For 56 farmers, we observe fresh weights but not dry weights. For 268 farmers, we are missing both fresh and dry weights.

Although farmers in the control group planted a wider range of seed varieties, the average expected time to maturity is comparable across treatment and control groups and aligns with our fresh weight data collection timing, so seed choice or program participation should not impact attrition due to missing harvest measurements. We drop a further 139 observations due to inconsistencies in outcome data or missing control variables.

In Appendix C, we include a more complete description of attrition from our sample and implement several imputation strategies to test the robustness of our results. We show that our results are consistent across various robustness checks, including missing-data imputation methods.

2.3.3 Sample description and balance tests

We report summary statistics and balance tests on our pre-specified control variables in Table 1, which confirms the importance of agriculture for this population: nearly eighty percent of participants earned more than half of their income from farm labor in the year prior to the study. On average, participants cultivated roughly one acre of maize, harvesting about half a ton per acre. Most participants are highly specialized in maize production, with 57% of participants cultivating maize exclusively at baseline. The median participant who cultivated any other crops did so on half an acre total at baseline. While three-quarters of the sample used improved agricultural technologies at baseline, average input intensity is low, especially for fertilizer. About half of the respondents report some knowledge of 1AF planting practices, which is primarily driven by the pre-exposed farmers.

Although household size, education, and baseline maize yields differ significantly across

Table 1: Baseline balance by treatment assignment, full sample

Variable	(1) Control Mean (SE)	(2) Treatment Mean (SE)	Difference (2)-(1)
Married (0/1)	0.88 (0.01)	0.88 (0.01)	0.01
Household head has secondary school (0/1)	0.38 (0.02)	0.44 (0.02)	0.05*
Household income >50% from farm labor (0/1)	0.78 (0.01)	0.78 (0.01)	0.00
Used improved ag technology in 2016 (0/1)	0.78 (0.01)	0.81 (0.01)	0.02
Reports knowledge of 1AF practices (0/1)	0.47 (0.02)	0.51 (0.02)	0.04
Intercropped maize and beans in 2016 (0/1)	0.48 (0.02)	0.48 (0.02)	0.00
Reports having credit access in 2016 (0/1)	0.71 (0.01)	0.74 (0.02)	0.03
Household size	6.66 (0.08)	6.90 (0.09)	0.26**
Acres under maize cultivation in 2016	1.01 (0.03)	1.05 (0.03)	0.04
Maize yield (kg/acre) in 2016	537.01 (13.52)	587.77 (16.25)	51.99**
F-statistic (test of joint significance)			0.82
Number of observations			1896

Note: field office area fixed effects included in all balance test regressions. Standard errors are clustered at the treatment assignment (farmer group cluster) level.

treatment and control, an F-test of joint orthogonality of the variables in Table 1 does not reject the null that all the variables are jointly orthogonal to treatment status. None of the three significantly different variables are highly correlated with our outcome variables, so the differences are unlikely to affect our results.[14] Nevertheless, we report results both with and without these control variables.

In addition to the above balance checks for treatment and control groups, we compare baseline characteristics of our two sub-populations: the "primary" sample (farmers who have never enrolled in 1AF) and the "pre-exposed," who had previously enrolled in 1AF programming. The two groups together make up the "full sample." These tables are shown in Appendix A. We observe imbalance in some individual variables but fail to reject the null that the variables are jointly orthogonal to treatment status.[15]

We also compare participants in the primary sample to those who enrolled prior to the study (Table A.3). At baseline, farmers who self-selected into early program access

[14] It may seem counter-intuitive that baseline maize yields do not correlate with maize yields, but these variables are very noisy and low autocorrelation is common. Baseline yields are also self-reported, in contrast to our endline measurement using crop-cuts. More broadly, self-reported yields often show very low correlation with crop-cut yields (see for example Lobell, Azzari, Burke, Gourlay, Jin, Kilic, and Murray (2020)).

[15] Our preferred specification does not include control variables. In Appendix C.2, we show two alternative specifications, one including the set of pre-specified control variables from the pre-analysis plan and a second including a much larger set of potential candidate variables which are then selected using a double-Lasso procedure.

are more educated, cultivate more maize land, report using improved inputs and having access to credit, and are more likely to have prior knowledge of 1AF practices. We cannot determine whether this is due to early enrollees being better off to start or whether program participation contributed to the observed differences.

2.3.4 Data collection and quality validation

The data collection was directly managed by 1AF, which has an independent data collection department. To assuage potential concerns about the independence of the research design, the research team took several steps. First, the International Initiative for Impact Evaluation (3ie) helped design and review all parts of the trial (the experimental design, field protocols, sampling, randomization, and data collection instruments (Dubey and Yegbemey, 2017)). Second, the data collection followed best-practice protocols for data collection, including back-checks at all phases of data collection. Third, 1AF contracted an independent firm, Intermedia Development Consultants, to carry out a three-step audit of the data collection. Fourth, as described above, maize yields were physically weighed and cultivated maize acreage was measured with GPS, respectively,

All weighing and surveying was carried out by enumerators hired by the 1AF Monitoring and Evaluation department. These were not program staff, came from outside the study area, and did not know the sample farmers. Finally, two of the authors on this paper were brought in as independent evaluators to insure independence. The non-1AF part of the team reviewed the pre-analysis plan (PAP) prior to data collection and independently conducted the data cleaning, variable construction, and analysis (Deutschmann and Tjernström, 2018). Appendix F provides more details on the auditing procedures.

3 Empirical strategy

We obtain intent-to-treat estimates of program impacts by estimating the following regression with OLS:

$$y_{ics} = \alpha + \beta T_{cs} + [\delta X_{ics}] + \gamma_s + \varepsilon_{ics} \tag{1}$$

where y_{ics} denotes the outcome for individual i in cluster c and strata s, with strata defined based on 1AF field-office areas. The indicator variable T_{cs} takes on a value of one for clusters assigned to treatment. X_{ics} is a vector of controls, γ_s is a strata fixed effect to account for variations in the probability of assignment to treatment, and ε_{ics} is clustered at the level of treatment assignment (farmer group clusters).

In the body of the paper, we report results which include a control for past exposure to the 1AF program where appropriate, but no other controls. We show results in Appendix C which include either our pre-specified set of controls or a set of controls selected via a double lasso procedure (Belloni, Chernozhukov, and Hansen, 2014; Urminsky, Hansen, and Chernozhukov, 2016). Our PAP specified X_{ics} in Eq. 1 as including controls for household demographics (marital status, household size, education, credit access, land ownership, agricultural reliance) and baseline agricultural characteristics (technology use, intercropping, knowledge of 1AF practices). Additionally, the PAP proposed to include a spillover inverse probability weight, with spillover likelihood measured by a farmer's total agricultural contacts that were randomized into treatment. Including this weighting does not change the statistical significance or qualitative interpretation of our results (results available upon request).

The standard errors reported in our main results are not adjusted for multiple hypothesis testing. The interpretation of our results does not change when we account for multiple hypothesis testing using Westfall and Young (1993) and Holm-Bonferroni methods, nor when we apply Fisher's exact test in a randomization inference procedure following Young (2019).

It is worth noting that, although we estimate an intent-to-treat specification, in practice our estimates are close to giving the average treatment effect on the treated due to high compliance (i.e., enrollment in 1AF). The experiment was designed to maximize power for detecting impacts on yields and profits, and so randomized among a set of highly likely compliers who had already completed a baseline survey, paid a deposit, and formed a farmer group to participate in the 1AF program. We do face attrition in some of our outcomes, as discussed above briefly in Section 2.3.2 and more fully in Appendix C.1, but our main estimates reported in the body of the paper should be interpreted with this high treatment assignment compliance in mind.

4 Results: maize yield and output

The primary goal of the 1AF small farmer program is to increase the maize productivity of participating farmers. We consider two related outcomes in this section: maize yields and total maize output. Maize yields are particularly important if we are concerned about allocative efficiency and structural transformation (Dercon and Gollin, 2014; Gollin, 2015). Understanding total maize output allows us to understand how the program affects overall farmer welfare, a topic we discuss more extensively below in Section 6.

Participation in the 1AF program has an economically and statistically significant effect on both maize yields and total maize output. In columns 1 and 2 of Table 2, we find that maize yields (kg per acre) increase by 26-27% in the treatment group, whether we consider the primary sample or the sample including pre-exposed farmers. As described above, for treated farmers we consider yields on the land farmers enrolled in the 1AF program. On average, enrolled land among the treatment group represents 63% of total cultivated maize acreage. We also observe maize production on the non-enrolled land in the treatment group. Note that the average fraction of land enrolled does not differ systematically by primary vs. pre-exposed. Pre-exposed farmers enroll more land in the program, but they also farm slightly more land overall, which results in a similar fraction of enrolled land. This is true both for actual measured enrolled plot size, and for the coarser admin data enrollment number, which measures packages in quarter-acre steps.

Next, we examine whether the increased yields on enrolled land comes at the expense of yields on non-enrolled land. Columns 3 and 4 of Table 2 compare yields on treated farmers' non-enrolled land to yields on control farmers' land. This analysis shows that farmers in the treatment group are not increasing yields on their enrolled land at the expense of yields on their non-enrolled land, as yields on non-enrolled land are statistically similar to yields among control farmers. This may be surprising, as we might expect farmer practices—or even their input use—to spill over to their non-enrolled plots. Additional analysis in Table C.7 reveals that treated farmers are indeed more likely to use 1AF-recommended practices on non-enrolled land relative to control farmers; however, these differences are relatively small in magnitude. Table C.8 shows that farmers are substantially more likely to adopt recommendations on enrolled land relative to non-enrolled land. A potential explanation for this limited information spillover could be that it is partly driven by 1AF's programming: practices are presented as a bundle, and farmers are encouraged to adopt all of them.

Another way of looking at productivity is to examine impacts on farmers' maize output overall. In columns 5 and 6 of Table 2, we can see that that farmers in the treatment group are not only more productive per acre of enrolled land, but they increase their overall maize output by 24% relative to control farmers. It appears that farmers are using their land more efficiently and harvesting more maize as a result.

Across the board, the treatment effects on yields and output are not statistically different between pre-exposed farmers in the treatment and control groups. The point estimates on the interaction between random assignment to the program and past participation are negative, suggesting the benefits may be slightly lower for those farmers, but the coefficients are consistently imprecisely estimated. Our interpretation of this is that participation in the 1AF program is beneficial for farmers beyond a single season—a finding that is per-

Table 2: Yield and output

	Maize yields enrolled vs. control		Maize yields non-enrolled vs. control		Total maize output	
	(1)	(2)	(3)	(4)	(5)	(6)
1AF participant	295.43***	306.82***	-52.15	-37.81	264.90***	278.78***
	(37.230)	(37.970)	(36.790)	(41.090)	(85.730)	(90.560)
Pre-exposed		20.05		24.28		192.43**
		(34.130)		(33.870)		(74.320)
1AF participant		-14.05		-10.32		-33.33
× pre-exposed		(49.290)		(55.560)		(120.710)
1AF + (*1AF* × *Pre-exposed*)		292.77***		-48.13		245.45***
		(30.37)		(40.47)		(66.92)
Control group mean	1128.39	1150.49	1128.39	1150.49	1082.38	1164.00
Observations	682	1896	614	1701	682	1896

Note: This table presents results from linear regressions of the outcomes at the top of each column on the treatment dummy. Maize yields is measured as kgs of maize per acre cultivated, and total maize output is measured in kgs. In columns 1 and 2, enrolled refers to the land farmers in the treatment group enrolled in the 1AF program, which may be less than the full amount of land on which they cultivate maize. In columns 3 and 4, we include only farmers who also cultivated maize on land they did not enroll with 1AF and compare the yields from that non-enrolled land to yields from control group farmers. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. Columns 1, 3, and 5 include only farmers who had never previously participated in the 1AF program. Columns 2, 4, and 6 additionally include a sample of farmers who had previously enrolled in the 1AF program (see Section 2.2). All regressions include strata (field office) fixed effects.

haps unsurprising from a revealed preference perspective, given that two-thirds of farmers re-enroll in the program the following year (see Appendix E for more details).

5 Mechanisms

The bundled nature of the intervention we study in this paper makes it difficult to fully disentangle the mechanisms driving the effects we observe. Previous work studying farmer constraints to technology adoption and productivity has typically focused on isolating single constraints. We provide suggestive evidence consistent with multiple constraints playing a role in farmers' decisions and maize production, suggesting the bundled program is effective at shifting farmers' constraints along multiple dimensions. Indeed, it is highly likely that there are important complementarities between these dimensions, and the effects of the bundled program may be greater than the sum of its parts.

5.1 Information and behavioral changes

We first examine several practice choices farmers make during production. These are primarily or purely informational constraints: farmers could adopt these practices even absent the 1AF program or the credit it provides for input purchases. If we observe treatment effects on adoption of these practices, it would be consistent with the program easing information constraints and inducing farmers to adopt improved practices.

Table 3 reports treatment effects on key practices, estimated with linear probability models. The outcome variable in columns 1 and 2 is a dummy equal to one if farmers planted their rows within 5 cm of 1AF's recommended spacing. Similarly, columns 3 and 4 reports regressions on a dummy equal to one if farmers spaced their plants within rows within 5cm of 1AF's recommended spacing. Enumerators measured this spacing at several points on farmers' fields, and our outcome variable is based on a comparison of the average of those measurements to the recommendations in 1AF's training guides.

High maize productivity requires substantial amounts of nitrogen—but the timing matters. Maize plants require relatively little nutrient input at the time of planting and are most responsive to inputs later in the season. Accordingly, farmers are recommended to apply DAP (diammonium phosphate) at planting and CAN (calcium ammonium nitrate) several weeks later. The outcome variable in columns 5 and 6, fertilizer timing, takes on a value of one if the participant applied fertilizer at the appropriate time of the season.

Across the board, the treatment farmers are much more likely to follow 1AF's recommended practices. Columns (1)-(4) show that treated participants are more likely than control farmers to follow row- and plant-spacing recommendations. Adherence to recommended spacing increases by almost 60% (220%) for row spacing (plant spacing) among farmers unacquainted with the program. We can see in columns (5) and (6) that treated farmers are 240% more likely to use appropriate fertilizer timing.

Based on what we know from the literature, we should not be surprused that, absent the program, even experienced farmers may overlook details like the proper fertilizer timing—even though this might seem like a basic productivity-enhancing step. Research suggests that even highly-experienced producers can fail to notice crucial features of the production process (Hanna, Mullainathan, and Schwartzstein, 2014) or less-salient profitability margins (Beaman et al., 2013; Duflo, Kremer, and Robinson, 2008). Fertilizer recommendations in the region typically focus on the amount of different types of fertilizer rather than on application timing. While extension manuals often distinguish between fertilizer application at planting and at top dressing, they rarely emphasize the crucial importance of timing (e.g. National Farmers Information Services 2019).

Table 3: Behavioral changes

	Row Spacing		Plant Spacing		Fertilizer Timing	
	(1)	(2)	(3)	(4)	(5)	(6)
1AF participant	0.22***	0.24***	0.20***	0.21***	0.65***	0.65***
	(0.040)	(0.040)	(0.030)	(0.030)	(0.030)	(0.030)
Pre-exposed		0.07**		0.02		0.15***
		(0.030)		(0.020)		(0.030)
1AF participant		-0.11**		-0.07*		-0.18***
× pre-exposed		(0.050)		(0.040)		(0.040)
1AF + (*1AF* × *Pre-exposed*)		0.13***		0.14***		0.47***
		(0.03)		(0.02)		(0.03)
Control group mean	0.37	0.41	0.09	0.13	0.27	0.38
Observations	682	1896	682	1896	682	1896

Note: This table presents results from linear regressions of the outcome variables in each column on the treatment dummy. Row and plant spacing are indicator variables for whether or not a farmer was within 5cm of the recommended spacing, for the rows in which they planted and the plants within rows, respectively. Fertilizer timing is an indicator variable for whether or not farmers applied DAP at planting and CAN at top dressing (the timing recommended by the NGO). Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. Columns 1, 3, and 5 include only farmers who had never previously participated in the 1AF program (the "primary" sample). Columns 2, 4, and 6 additionally include a sample of farmers who had previously enrolled in the 1AF program (the "full" sample; see Section 2.2). All regressions include strata (field office) fixed effects.

Table 3 also reports estimated coefficients on previous program exposure and its interaction with the treatment. The results show that the treatment effects on 1AF-promoted practices are substantially larger for new participants. Nevertheless, pre-exposed farmers also experience positive and significant treatment effects along all three dimensions. It appears that even seemingly basic information about practices may bear repeating multiple times, albeit with returns that may diminish over time. Table A.4 explores this a bit further, by showing the relationship between baseline knowledge and adoption of recommended practices in the primary sample. Looking at only control-group participants (columns (1), (3), and (5)), we find no statistically significant relationship between baseline knowledge of 1AF recommendations and current-season adoption of these practices. Columns (2), (4), and (6) add treatment-group participants and again find no evidence of heterogeneous treatment effects by baseline knowledge.

5.2 Liquidity constraints and investment

Input intensification is an important contributor to agricultural productivity gains. Table 4 therefore examines program impacts on the intensive margins of production. The main variables are expenditure on fertilizer, seeds, labor (both paid and unpaid, with the latter valued as described in section 2.3), and land. Enrolled farmers in the primary sample spend 91% more on fertilizer, 26% more on seeds, 18% more on paid labor, 19% more on unpaid labor, and dedicate 13% more land to maize. Farmers in treated and control groups report using a similar fraction of total available land for maize cultivation, with the median farmer reporting 83% of total cultivated acreage is dedicated to maize production. These are substantial expenditure increases, which we account for in our profit analysis below. It is worth noting that 1AF's average gross margin on inputs is 32%, which is similar to markups in the agro-dealer sector in the region (Tinsley and Agapitova, 2018). Therefore, we do not expect that these cost increases stem from 1AF's input prices.

Compared to the above results on behavioral changes, which suggest more muted treatment effects of program-provided information on pre-exposed farmers, we find no differential treatment effects on farming investments for the pre-exposed group. Pre-exposed farmers in the treatment group increase their investment along all the investment dimensions that we consider by a magnitude that is comparable to the primary sample. If information were the primary constraint preventing farmers from using appropriate quantities of seeds and fertilizer, we might expect the treatment effect on these margins to decrease over time as farmers learn, similar to the effects we observe in Table 3. If farmers are only marginally credit constrained, we might also expect the credit from the 1AF program to permit in-

Table 4: Use of costly inputs

	Fertilizer		Seeds		Paid Labor		Unpaid Labor		Maize Acres	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1AF participant	18.66***	18.88***	3.84***	4.01***	5.29*	5.83*	3.36***	3.59***	0.12*	0.12**
	(1.940)	(1.910)	(0.810)	(0.820)	(2.930)	(3.030)	(1.160)	(1.250)	(0.060)	(0.060)
Pre-exposed		5.67***		1.23		6.13***		-1.17		0.14***
		(1.730)		(0.750)		(1.900)		(1.140)		(0.050)
1AF participant		0.47		0.62		2.95		1.59		-0.03
× pre-exposed		(2.620)		(1.140)		(3.700)		(1.450)		(0.080)
1AF + (*1AF* × *Pre-exposed*)		19.35***		4.63***		8.78***		5.18***		0.10**
		(1.52)		(0.71)		(2.15)		(0.93)		(0.04)
Control group mean	20.58	25.30	14.77	15.59	29.65	31.90	17.83	16.66	0.93	0.98
Observations	682	1896	682	1896	682	1896	682	1896	682	1896

Note: This table presents results from linear regressions of the outcome variables in each column on the treatment dummy. Costs are expressed in USD. For more on how we define labor costs, see Appendix B. Columns 1-8 show input costs in USD. Columns 9 and 10 show all acres used for maize cultivation, including both land enrolled in the 1AF program and non-enrolled land. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. Odd columns include only farmers who had never previously participated in the 1AF program. Even columns additionally include a sample of farmers who had previously enrolled in the 1AF program (see Section 2.2). All regressions include strata (field office) fixed effects.

creased investments, increased profits, and subsequently similarly reduced effects on these dimensions. We interpret these results as instead suggesting that farmers are substantially credit constrained, even after a year or more of program participation. In other words, the credit and potential profit from one more year of program participation do not fully resolve the constraints they face.

As further evidence that credit constraints play a key role in farmers' decisions, we now turn to major categories of household expenditures reported by farmers before and after treatment assignment. These all cover the period *before* farmers received any inputs from 1AF. Changes in expenditures here therefore represent responses to an upcoming relaxation of credit constraints. In Table 5, we see that farmers in the treated group increase expenditures on school fees by 50-70% relative to the control group following treatment assignment. This might suggest that treatment farmers were able to allocate limited savings, which had perhaps been put aside for input purchases, towards school fees in the short-run. Overall, educational expenditures among pre-exposed farmers are higher than in the primary sample, but the treatment effect does not differ significantly across the two groups. This is consistent with a progressive relaxation of credit constraints with repeated program participation—but could also simply be based on selection, such that pre-exposed farmers are better off to start. We see no detectable changes in medical, celebration, or livestock expenditures. School fees constitute a relatively large fraction of reported household expenditures compared to these other categories, so it may be that a relaxation of credit constraints is relatively less important for those expenditure categories, or that those expenditures can more easily be spread out or incurred during different times than the planting season.

Table 5: Major expenditures, 3 months after treatment assignment

	Education		Medical		Celebration		Livestock	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1AF participant	55.78***	54.97**	-0.57	-0.78	3.23	3.60	-1.08	-0.63
	(18.990)	(21.950)	(0.500)	(0.610)	(3.850)	(4.460)	(3.110)	(3.060)
Pre-exposed		56.33***		0.31		0.23		2.90
		(17.090)		(1.160)		(1.530)		(2.450)
1AF participant		-15.77		1.23		-3.84		-2.01
× pre-exposed		(31.680)		(1.770)		(4.590)		(3.740)
1AF + (*1AF* × *Pre-exposed*)		39.21**		0.45		-0.24		-2.64
		(19.03)		(1.61)		(1.71)		(2.17)
Control group mean	79.48	110.22	1.09	1.17	1.34	1.64	7.39	8.17
Observations	682	1896	682	1896	682	1896	682	1896

Note: This table presents results from linear regressions of the outcome variables in each column on the treatment dummy. The outcomes in this table show the respondent's expected expenditures for the three months after treatment assignment. Expenditures are in USD. Each regression controls for reported expenditures. in the same category for the three months before treatment assignment. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. Odd columns include only farmers who had never previously participated in the 1AF program. Even columns additionally include a sample of farmers who had previously enrolled in the 1AF program (see Section 2.2). All regressions include strata (field office) fixed effects.

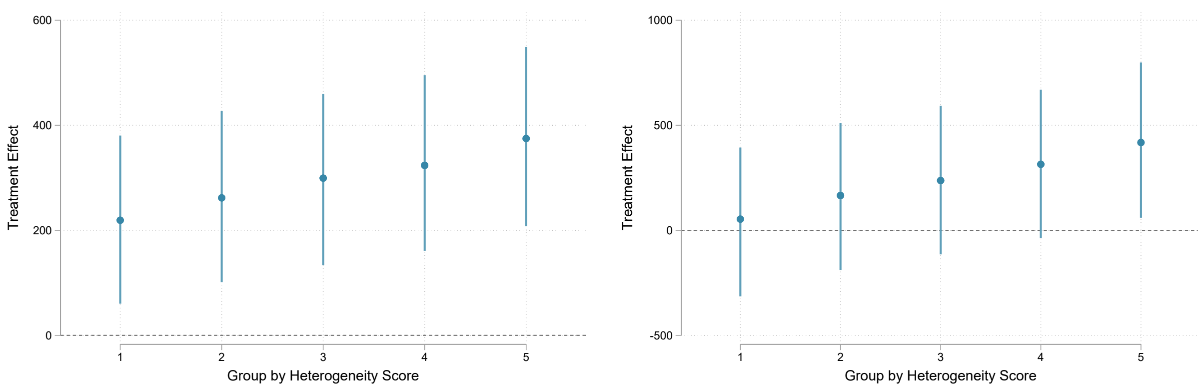
5.3 Heterogeneity

Heterogeneity analysis may allow us to further explore the potential mechanisms driving the average effects we observe above. For example, we observe baseline indicators of self-reported credit access. If program effects were greater for farmers without credit access, we might conclude that the credit component of the program is particularly important. To explore this in an agnostic way about which covariates may be predictive, we implement machine learning methods developed by Chernozhukov, Demirer, Duflo, and Fernandez-Val (2018).

Figure 2 shows the estimated Group Average Treatment Effects that result from applying the methods of Chernozhukov et al. (2018). We observe that treatment effects on both maize yields and total maize output show little evidence of heterogeneity between groups, although Figure 2b suggests that effects on total output may be attenuated for some farmers. The estimation procedure also produces what Chernozhukov et al. (2018) term the 'heterogeneity loading' parameter of the Best Linear Predictor of the conditional average treatment effect. For all outcomes and methods, we fail to reject the null hypothesis that this heterogeneity parameter is statistically different from zero.

To further corroborate these results, we conduct a Monte Carlo simulation exercise of the type suggested in Appendix E of Heckman, Smith, and Clements (1997). To simulate the distribution of impact standard deviations under the null hypothesis of no heterogeneity, we repeatedly sample the control group to generate synthetic treatment and control groups. This gives us a distribution of the standard deviation of percentile effect differences under the null, which we then compare to the impact standard deviation seen in the data.

For all outcomes, we fail to reject the null of no heterogeneity, suggesting that we are unable to detect treatment effect heterogeneity under the assumption of perfect positive dependence between treatment and control outcome percentiles (also called the location shift assumption). Since this is a strong assumption, we also implement the rank preservation test proposed in Bitler, Gelbach, and Hoynes (2005). This tests for rank reversal in baseline characteristics between quartiles of the treatment and control distribution. For each sample definition and outcome variable, we fail to reject the null in the test for joint-orthogonality.[16] These exercises together suggest limited scope to find major heterogeneity in treatment effects. In other words, the program seems broadly effective at relieving multiple constraints for farmers, and there does not seem to be one group which is relatively unaffected by this bundled intervention.



(a) Maize Yields

(b) Total Maize Output

Figure 2: Sorted Group Average Treatment Effects, estimated with Neural Nets

[16] Results of both exercises are available from the authors upon request.

6 Welfare and Net Benefits

We have shown that participation in the 1AF program increases farmers' maize production, but that participating farmers also increase their expenditures on seeds, fertilizer, and labor. To determine the net effect on farmer welfare, we construct an estimate of farmer maize profits that accounts for the market value of maize production. In Table 6, we estimate that farmer profits in the treatment group increase by $56-63 USD, or 17-18%. This finding is robust to a variety of assumptions about the value of unpaid labor. Our pre-analysis plan specified valuing unpaid labor at 50% of the day wage in local markets. Recent work by Agness et al. (2022) in a similar setting suggests valuing unpaid labor at 60% of the market wage. Even conservatively assigning the full day wage to unpaid work results in qualitatively similar conclusions about profits.

One possible concern regarding labor use is that farmers who increase labor use on maize land reduce labor allocations on other crops, reducing the overall welfare impacts of program participation. We do not observe labor use on land allocated for other crops, which limits our ability to rule out such substitution effects. We observe that treated farmers increase unpaid labor person-days by 25%, but in levels, this means that unpaid labor use increases by less than one day per household member in the median household, suggesting any change in unpaid labor availability for other crops was likely to be relatively minimal.

The effect of program participation on profits is attenuated among pre-exposed farmers in the treatment group. Farmers in the pre-exposed group are farming more profitably in general. We cannot say with certainty whether increased profits among pre-exposed farmers represent past effects of program participation or selection into the program in the past. However, given our observations above that some benefits of participation are attenuated among the pre-exposed sample, it is plausible that for some of these farmers the benefits of participating in the full program no longer outweigh the costs. Farmers may nevertheless find it worthwhile given credit constraints and the quality of inputs available through the 1AF program.

6.1 Net benefits under different assumptions

Beyond the direct effect on farmer profits of program participation, we also consider the broader question of whether the benefits that we have estimated above are worth the overall costs of 1AF's intervention. Specifically, we conduct a simple cost-benefit analysis to determine the net economic impact of the 1AF program in Kenya given our estimated

Table 6: Profits

	Profit (PAP Definition)		Profit (Full Labor Costs)	
	(1)	(2)	(3)	(4)
1AF participant	59.67*	62.89*	56.31*	59.30*
	(30.210)	(32.470)	(30.050)	(32.180)
Pre-exposed		61.70**		62.87**
		(27.430)		(27.020)
1AF participant		-20.56		-22.15
× pre-exposed		(43.140)		(42.710)
1AF + (*1AF* × *Pre-exposed*)		42.33*		37.15
		(24.19)		(23.91)
Control group mean	332.11	361.58	314.29	344.92
Observations	682	1896	682	1896

Note: This table presents results from linear regressions of the outcomes shown at the top of each column on the treatment dummy. Profits are calculated as the value of maize production less costs (see Appendix B), and are reported in USD. Columns 1 and 2 follow the PAP and account for unpaid labor costs by valuing unpaid time at 50% of the average day-wage reported in the sample for paid labor. Columns 3 and 4 instead value unpaid labor at the full average day-wage reported in the sample. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. Odd columns include only farmers who had never previously participated in the 1AF program. Even columns additionally include a sample of farmers who had previously enrolled in the 1AF program (see Section 2.2). All regressions include strata (field office) fixed effects.

treatment effects. In international aid discussions, the concept of value for money is often equated with return on investment (ROI) metrics. That approach more closely resembles a financial analysis than an economic cost-benefit analysis; in particular, the latter aims to capture the net contribution that an intervention makes to social welfare.[17]

Rather than rely on single treatment effect point estimates, we carry out Monte Carlo simulations to account for different types of uncertainty. Our main sources of uncertainty are parameter uncertainty (due to the statistical uncertainty around our estimated treatment effects) and what the cost-benefit literature calls *structural uncertainty* (due to the assumptions made during analysis). We already explore the robustness of our main results to several sources of structural uncertainty in Appendix C, but here we look instead at the effect on net benefits.

The ideal approach would be to compute the discounted stream of expected utility from the program and subtract the discounted stream of costs, to arrive at the net present value (NPV) of the program. Instead of utility, we use our treatment effect estimate on profits as the main benefit. We would prefer to have consumption data instead of profits,

[17] To illustrate the difference, a financial analysis does not consider the influence of market distortions on project impacts or costs.

Table 7: Monte Carlo simulation parameters and probability distributions

Parameter	Distribution	Source
Donor subsidy	$\mathcal{U}(13.3, 37.3)$	1AF administrative data, 2016-2021[a]
Profit, new farmers	$\mathcal{N}(61.3,\ 32.5)$	Table 6, column (2)[b]
Profit, pre-exposed farmer	$\mathcal{N}(41.1,\ 32.5)$	Table 6, column (2)[b]

Note:
[a] We obtained internal data from 1AF on the average donor subsidy per enrolled farmer in Kenya for the period 2016-2021. While we cannot share the exact numbers, we use a uniform distribution parameterized with min and max values two standard deviations from the mean of the data points. In this table, we report the resulting minimum and maximum values. As an example, a One Acre Fund impact report cites an overall average donor subsidy of $26 across all country programs in 2016 (One Acre Fund, 2016).
[b] For simplicity, we assume that the distributions of profit treatment effects have the same standard deviation for new and pre-exposed farmers. We conservatively use the larger regression standard error as our estimate of the standard deviation.

as it more closely maps into utility; however, we did not collect consumption data as the study's primary focus was maize yields and agricultural profits. Given imperfect credit markets in this context, profits may be a reasonable proxy for consumption. The result in Table 5, which shows that program participants spend around $55 more on education—a number similar in magnitude to the average profit treatment effect—lends further support to our assumption that participants consume most of the increased income.

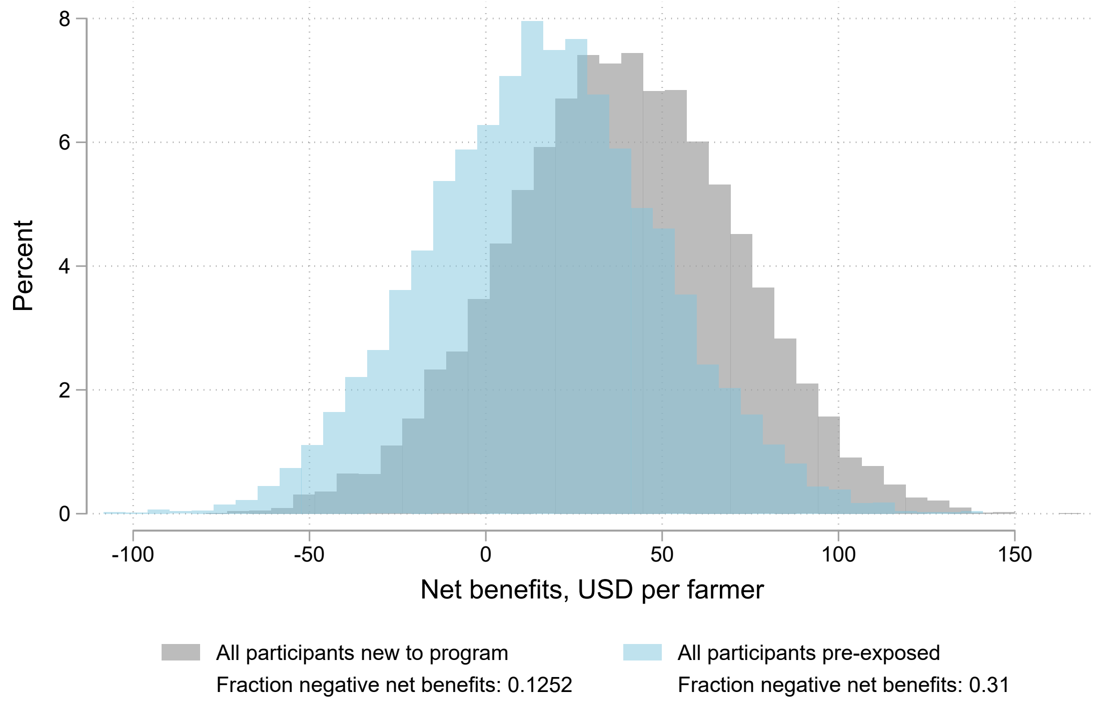
Table 7 details the probability distributions that we assume for this probabilistic analysis, and Figure 3 shows the net benefit distributions that emerge from the simulation exercise. The costs of administering the program vary some amount from year to year. To account for this, we use administrative data on the average per-farmer donor subsidy from 1AF's Kenya program spanning the time period 2016-2021.[18] We approximate the variation using a uniform distribution since the year-to-year variation is relatively small and since we don't have strong reasons to believe that any of the data points are more likely than others.

We first look at the scenario where all participants are new to the program. Drawing per-farmer donor subsidies and average profits across 10,000 trials, we derive the distribution of net benefits shown in Figure 3 by a gray histogram. This distribution of net benefits has a mean of 38.1 and a standard deviation of 33.1. Further, we estimate that the average net benefits of the program are positive in about 87 percent of cases.

To evaluate the net benefits of a program operating at scale, it may be more plausible to posit that all or many participants will already have been exposed to the program—much like in our study area. We estimate a second scenario in which we assume all participants

[18] This estimated per-farmer donor subsidy accounts for program overhead, including international staff, and R&D expenditures.

Figure 3: Distribution of net benefits under different assumptions



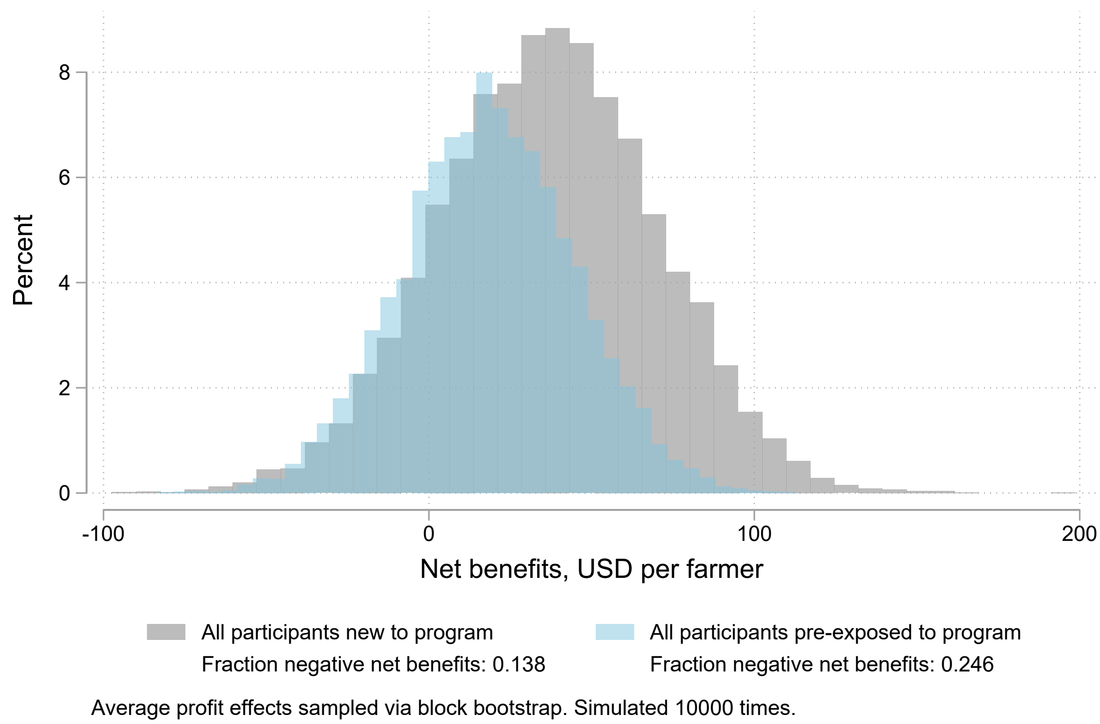
Note: This figure shows the distribution of net benefits under 2 scenarios. The gray distribution shows estimated net benefits when we assume all participants are new to the program. The blue distribution instead shows estimated net benefits when we assume all participants are similar to our pre-exposed sample.

instead experience benefits similar to the pre-exposed farmers in our sample. The light blue histogram in Figure 3 shows the resulting distribution of estimated net benefits. The mean and standard deviation of this distribution are slightly lower (16.4 and 33.1), with positive net benefits of program participation positive for about 69 percent of draws.

We additionally conduct simulations in which, rather than drawing estimated profits from a triangular distribution parameterized by our treatment effects, we use a block bootstrap to directly generate a distribution of profit impacts. As with the previous simulations, we consider two scenarios, comparing profit impacts on new and pre-exposed to estimates of donor subsidies. Figure 4 plots the results of this exercise, which suggests that the program generates a mean net benefit for new farmers of 36.6, with a standard deviation of 33.7 (for pre-exposed farmers, these numbers are 18 and 26.4, respectively); further, the net social benefits are positive in 86 and 75 percent of cases, for new and pre-exposed farmers, respectively.

In summary, our simulations suggest that the net social benefit of the program is positive—not only for participants themselves, but taking into account the fraction of

Figure 4: Distribution of net benefits under different assumptions, with profit impacts drawn by block bootstrap



Note: This figure shows the distribution of net benefits under 2 scenarios. The gray distribution shows estimated net benefits when we assume all participants are new to the program. The blue distribution instead shows estimated net benefits when we assume all participants are similar to our pre-exposed sample.

program fees that are not covered already by farmer fees.[19] Moreover, given that we find no evidence of heterogeneous treatment effects above in Section 5.3, we expect that the net benefit of participation should be broadly positive for participants at scale.

7 Conclusion

We uncover large ATEs from participation in 1AF's core program. The treatment effects are remarkably stable across sample definitions and specifications and we find limited evidence of heterogeneous effects. Our results are consistent with the large theoretical literature

[19]For the sake of comparison, we could estimate a rate of return for the implicit donor subsidy of the program, following Duflo et al. (2008). At an average donor subsidy of 25.3, and given the average increase in profit in the primary sample of 59.5, we would estimate a rate of return of 135.18. Another alternative would be to estimate a benefit-cost ratio similar to the calculations in Carter et al. (2021), which would yield a benefit-cost ratio accounting for the donor subsidy of 2.35, roughly 30 percent greater than that study's estimated benefit-cost ratio for direct benefits.

on poverty traps as well as empirical work showing that poor households benefit from bundled interventions (Balboni et al., 2022; Bandiera et al., 2017; Banerjee et al., 2015). If we believe that farmers face multiple simultaneous constraints, an intervention that only provides credit or information may not suffice to raise yields and profits in a meaningful way.

While we cannot answer the question of whether a bundled approach is more cost effective than a simpler or more targeted program, our work suggests some considerations that future work on this topic could explore further. A standardized program that targets multiple constraints may allow an organization to scale across space without tweaking program specifics in each new context. The cost-effectiveness of bundled programs therefore hinges on the costs associated with "over-bundling" relative to the cost of the market research required to tailor programs to each new context.

Returning to the broader question of the role of the agricultural sector in Africa's future and the contribution of smallholder farmers to economic growth, we recognize the limits of a single study in a particular region in a specific year. A deeper discussion of external validity (even as it applies to extrapolating to other 1AF locations) would be valuable, but is beyond the scope of the current paper. Nevertheless, the 1AF model addresses several of the concerns about smallholder farming systems that Collier and Dercon (2014) raise.

A primary reason why small farmers may struggle to compete relates to how scale influences capital costs, logistical capacity, and bargaining power. Large organizations can and do leverage grant funding for working capital from both agro-input suppliers and banks (Tinsley and Agapitova, 2018). Most financial institutions shy away from input loans, preferring to provide credit over shorter time periods. By extending their credit to smallholders, organizations like 1AF can effectively reduce the transaction costs and asymmetric information facing small farmers.

1AF also imports their own inputs, conducts their own quality controls, and manages their own storage facilities. In so doing, they have successfully integrated several parts of the value chain, allowing them to benefit from economies of scale—the benefits of which they can pass on to their clients. By integrating their supply chain, 1AF may also be able to offer better quality inputs to farmers on average than other shops in the market. Investigating this mechanism would be a fruitful direction for future research. In addition, a large organization like 1AF can leverage their scale and human capital to process and distill large amounts of information on agricultural productivity, risks, and complementary practices to share with small farmers when and where they need it. This allows small farmers to benefit from skills and technology transfers otherwise often out of reach for all but the largest market participants.

Perhaps this new evidence will nudge a few cynics into reconsidering the future for (smallholder) agriculture as low-income countries grow and develop. For those already optimistic about the sector, we hope that it provides compelling input into a discussion about optimal instruments for boosting productivity. Thinking about large-scale organizations like 1AF as enabling creative vertically-integrated opportunities for farmers may hold a clue to the types of investments needed to transform African agriculture into a more dynamic sector.

References

Abay, K. A., L. E. M. Bevis, and C. B. Barrett (2021): "Measurement Error Mechanisms Matter: Agricultural Intensification with Farmer Misperceptions and Misreporting," *American Journal of Agricultural Economics*, 103, 498–522.

Agness, D., T. Baseler, S. Chassang, P. Dupas, and E. Snowberg (2022): "Valuing the Time of the Self-Employed," Working Paper 29752, National Bureau of Economic Research.

Arouna, A., J. D. Michler, and J. C. Lokossou (2021): "Contract Farming and Rural Transformation: Evidence from a Field Experiment in Benin," *Journal of Development Economics*, 151, 102626.

Ashraf, N., J. Berry, and J. M. Shapiro (2010): "Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia," *American Economic Review*, 100, 2383–2413.

Balboni, C., O. Bandiera, R. Burgess, M. Ghatak, and A. Heil (2022): "Why Do People Stay Poor?*," *The Quarterly Journal of Economics*, 137, 785–844.

Bandiera, O., R. Burgess, N. Das, S. Gulesci, I. Rasul, and M. Sulaiman (2017): "Labor Markets and Poverty in Village Economies," *The Quarterly Journal of Economics*, 132, 811–870.

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Pariente, J. Shapiro, B. Thuysbaert, and C. Udry (2015): "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries," *Science*, 348, 1260799–1260799.

Beaman, L., D. Karlan, B. Thuysbaert, and C. Udry (2013): "Profitability of Fertilizer: Experimental Evidence from Female Rice Farmers in Mali," *American Economic Review: Papers and Proceedings*, 103, 381–86.

——— (2023): "Selection Into Credit Markets: Evidence From Agriculture in Mali," *Econometrica*, 91, 1595–1627.

Belloni, A., V. Chernozhukov, and C. Hansen (2014): "Inference on Treatment Effects after Selection among High-Dimensional Controls," *The Review of Economic Studies*, 81, 608–650.

Birkhaeuser, D., R. E. Evenson, and G. Feder (1991): "The Economic Impact of Agricultural Extension: A Review," *Economic Development and Cultural Change*, 39, 607–650.

Bitler, M. P., J. B. Gelbach, and H. W. Hoynes (2005): "Distributional Impacts of the Self-Sufficiency Project," Working Paper 11626, National Bureau of Economic Research.

Block, S. (2014): "The Decline and Rise of Agricultural Productivity in Sub-Saharan Africa since 1961," in *African Successes*, University of Chicago Press, for the National Bureau of Economic Research, 13–67.

Bold, T., S. Ghisolfi, F. Nsonzi, and J. Svensson (2022): "Market Access and Quality Upgrading: Evidence from Four Field Experiments," *American Economic Review*, 112, 2518–2552.

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

Bravo-Ortega, C. and D. Lederman (2005): "Agriculture and National Welfare around the World: Causality and International Heterogeneity since 1960," Working paper 3499, The World Bank.

Burke, M., L. F. Bergquist, and E. Miguel (2019): "Sell Low and Buy High: Arbitrage and Local Price Effects in Kenyan Markets*," *The Quarterly Journal of Economics*, 134, 785–842.

Carletto, C., S. Savastano, and A. Zezza (2013): "Fact or Artifact: The Impact of Measurement Errors on the Farm Size–Productivity Relationship," *Journal of Development Economics*, 103, 254–261.

Carter, M., R. Laajaj, and D. Yang (2021): "Subsidies and the African Green Revolution: Direct Effects and Social Network Spillovers of Randomized Input Subsidies in Mozambique," *American Economic Journal: Applied Economics*, 13, w26208.

Carter, M. R., E. Tjernström, and P. Toledo (2019): "Heterogeneous Impact Dynamics of a Rural Business Development Program in Nicaragua," *Journal of Development Economics*, 138, 77–98.

Chambers, R. (1983): *Rural Development: Putting the Last First*, Routledge.

Channa, H., J. Ricker-Gilbert, S. Feleke, and T. Abdoulaye (2022): “Overcoming Smallholder Farmers’ Post-Harvest Constraints through Harvest Loans and Storage Technology: Insights from a Randomized Controlled Trial in Tanzania,” *Journal of Development Economics*, 157, 102851.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018): “Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments,” Working Paper 24678, National Bureau of Economic Research.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013): “Inference on Counterfactual Distributions,” *Econometrica*, 81, 2205–2268.

Cohen, J. and P. Dupas (2010): “Free Distribution or Cost Sharing: Evidence from a Randomized Malaria Prevention Experiment,” *The Quarterly Journal of Economics*, 125, 1–45.

Cole, S. A. and A. N. Fernando (2021): “‘Mobile’Izing Agricultural Advice Technology Adoption Diffusion and Sustainability,” *The Economic Journal*, 131, 192–219.

Collier, P. and S. Dercon (2014): “African Agriculture in 50 Years: Smallholders in a Rapidly Changing World?” *World Development*, 63, 92–101.

de Janvry, A. and E. Sadoulet (2010): “Agricultural Growth and Poverty Reduction,” *World Bank Research Observer*, 25, 1–20.

Dercon, S. and D. Gollin (2014): “Agriculture in African Development: Theories and Strategies,” *Annual Review of Resource Economics*, 6, 471–492.

Desiere, S. and D. Jolliffe (2018): “Land Productivity and Plot Size: Is Measurement Error Driving the Inverse Relationship?” *Journal of Development Economics*, 130, 84–98.

Deutschmann, J. W., T. Bernard, and O. Yameogo (2023): “Contracting and Quality Upgrading: Evidence from an Experiment in Senegal,” Working Paper.

Deutschmann, J. W. and E. Tjernström (2018): “The Impact of One Acre Fund’s Small Farm Program,” Technical Report.

Diop, B. Z. (2024): “Upgrade or Migrate: The Consequences of Fertilizer Subsidies on Household Labor Allocation,” Working Paper.

Dubey, P. and R. N. Yegbemey (2017): "Technical Support to the Impact Evaluation of the Core One Acre Fund Program on Yields and Profits of Maize Farmers in Teso, Kenya," Field Report, International Initiative for Impact Evaluation.

Duflo, E., M. Kremer, and J. Robinson (2008): "How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya," *American Economic Review: Papers and Proceedings*, 98, 482–488.

Feder, G., R. E. Just, and D. Zilberman (1985): "Adoption of Agricultural Innovations in Developing Countries: A Survey," *Economic Development and Cultural Change*, 33, 255–298.

Field, E., R. Pande, J. Papp, and N. Rigol (2013): "Does the Classic Microfinance Model Discourage Entrepreneurship among the Poor? Experimental Evidence from India," *American Economic Review*, 103, 2196–2226.

Fishman, R., S. C. Smith, V. Bobić, and M. Sulaiman (2022): "Can Agricultural Extension and Input Support Be Discontinued? Evidence from a Randomized Phaseout in Uganda," *The Review of Economics and Statistics*, 104, 1273–1288.

Gollin, D. (2015): "Agriculture as an Engine of Growth and Poverty Reduction," in *Economic Growth and Poverty Reduction in Sub-Saharan Africa*, ed. by A. McKay and E. Thorbecke, Oxford University Press, 91–121.

Gollin, D., C. W. Hansen, and A. M. Wingender (2021): "Two Blades of Grass: The Impact of the Green Revolution," *Journal of Political Economy*, 129, 2344–2384.

Gourlay, S., T. Kilic, and D. B. Lobell (2019): "A New Spin on an Old Debate: Errors in Farmer-Reported Production and Their Implications for Inverse Scale - Productivity Relationship in Uganda," *Journal of Development Economics*, 141, 102376.

Government of Kenya (2010): "Agricultural Sector Development Strategy 2010-2020," Technical Report.

Hanna, R., S. Mullainathan, and J. Schwartzstein (2014): "Learning Through Noticing: Theory and Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 129, 1311–1353.

Heckman, J. J., J. Smith, and N. Clements (1997): "Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts," *The Review of Economic Studies*, 64, 487–535.

Hsu, E. and A. Wambugu (2022): "Can Informed Buyers Improve Goods Quality? Experimental Evidence from Crop Seeds," Working Paper.

Jayne, T., J. Chamberlin, L. Traub, N. Sitko, M. Muyanga, F. K. Yeboah, W. Anseeuw, A. Chapoto, A. Wineman, C. Nkonde, and R. Kachule (2016): "Africa's Changing Farm Size Distribution Patterns: The Rise of Medium-Scale Farms," *Agricultural Economics*, 47, 197–214.

Jayne, T. and S. Rashid (2013): "Input subsidy programs in sub-Saharan Africa: a synthesis of recent evidence," *Agricultural Economics*, 44, 547–562.

Jayne, T. S., J. Chamberlin, and R. Benfica (2018a): "Africa's Unfolding Economic Transformation," *The Journal of Development Studies*, 54, 777–787.

Jayne, T. S., N. M. Mason, W. J. Burke, and J. Ariga (2018b): "Review: Taking stock of Africa's second-generation agricultural input subsidy programs," *Food Policy*, 75, 1–14.

Kirimi, L., N. Sitko, T. S. Jayne, F. Karin, M. Muyanga, M. Sheahan, J. Flock, and G. Bor (2011): "A Farm Gate-to-Consumer Value Chain Analysis of Kenya's Maize Marketing System," Technical Report WPS 44/2011, Tegemeo Institute of Agricultural Policy and Development.

Ligon, E. A. and E. Sadoulet (2011): "Estimating the Effects of Aggregate Agricultural Growth on the Distribution of Expenditures," Working paper 1115, Department of Agricultural & Resource Economics, UC Berkeley.

Lobell, D. B., G. Azzari, M. Burke, S. Gourlay, Z. Jin, T. Kilic, and S. Murray (2020): "Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis," *American Journal of Agricultural Economics*, 102, 202–219.

Lowder, S. K., J. Skoet, and T. Raney (2016): "The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide," *World Development*, 87, 16–29.

Macchiavello, R. and J. Miquel-Florensa (2019): "Buyer-Driven Upgrading in GVCs: The Sustainable Quality Program in Colombia," Working Paper.

Magruder, J. R. (2018): “An Assessment of Experimental Evidence on Agricultural Technology Adoption in Developing Countries,” *Annual Review of Resource Economics*, 10, 299–316.

McArthur, J. W. and G. C. McCord (2017): “Fertilizing Growth: Agricultural Inputs and Their Effects in Economic Development,” *Journal of Development Economics*, 127, 133–152.

Michelson, H., A. Fairbairn, B. Ellison, A. Maertens, and V. Manyong (2021): “Misperceived Quality: Fertilizer in Tanzania,” *Journal of Development Economics*, 148, 102579.

National Farmers Information Services (2019): “Field Management – NAFIS,” Report, http://www.nafis.go.ke/agriculture/maize/field-management-practices/.

One Acre Fund (2016): “Comprehensive Impact Report,” Report, https://oneacrefund.org/publications/comprehensive-impact-report.

——— (2020): “How We Grow,” Report, https://oneacrefund.org/what-we-do/how-we-grow/.

Park, S., Z. Yuan, and H. Zhang (2023): “Technology Training, Buyer-Supplier Relationship, and Quality Upgrading in an Agricultural Supply Chain,” *The Review of Economics and Statistics*, 1–46.

Ravallion, M. and S. Chen (2007): “China’s (Uneven) Progress against Poverty,” *Journal of Development Economics*, 82, 1–42.

Rosenzweig, M. R. and C. Udry (2020): “External Validity in a Stochastic World: Evidence from Low-Income Countries,” *The Review of Economic Studies*, 87, 343–381.

Sheahan, M., R. Black, and T. S. Jayne (2013): “Are Kenyan Farmers Under-Utilizing Fertilizer? Implications for Input Intensification Strategies and Research,” *Food Policy*, 41, 39–52.

Suri, T. (2011): “Selection and Comparative Advantage in Technology Adoption,” *Econometrica*, 79, 159–209.

Suri, T. and C. Udry (2022): “Agricultural Technology in Africa,” *Journal of Economic Perspectives*, 36, 33–56.

Tamim, A., A. Harou, M. Burke, D. Lobell, M. Madajewicz, C. Magomba, H. Michelson, C. Palm, and J. Xue (2024): "Relaxing Credit and Information Constraints: Five-Year Experimental Evidence from Tanzanian Agriculture," *Economic Development and Cultural Change*, publisher: The University of Chicago Press.

Tarozzi, A., A. Mahajan, B. Blackburn, D. Kopf, L. Krishnan, and J. Yoong (2014): "Micro-Loans, Insecticide-Treated Bednets, and Malaria: Evidence from a Randomized Controlled Trial in Orissa, India," *American Economic Review*, 104, 1909–1941.

Tinsley, E. and N. Agapitova (2018): "Private Sector Solutions to Helping Smallholders Succeed: Social Entreprise Business Models in the Agriculture Sector," Report, The World Bank.

Tjernström, E., M. R. Carter, and T. Lybbert (2018): "The Dirt on Dirt: Soil Characteristics and Variable Fertilizer Returns in Kenyan Maize Systems," Working Paper.

Udry, C., F. di Battista, M. Fosu, M. Goldstein, A. Gurbuz, D. Karlan, and S. Kolavalli (2019): "Information, Market Access and Risk: Addressing Constraints to Agricultural Transformation in Northern Ghana," Draft Report.

Urminsky, O., C. Hansen, and V. Chernozhukov (2016): "Using Double-Lasso Regression for Principled Variable Selection," Working Paper.

Westfall, P. H. and S. S. Young (1993): *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*, John Wiley & Sons, Inc.

Wollburg, P., T. Bentze, Y. Lu, C. Udry, and D. Gollin (2024): "Crop Yields Fail to Rise in Smallholder Farming Systems in Sub-Saharan Africa," *Proceedings of the National Academy of Sciences*, 121, e2312519121.

World Bank (2008): "World Development Report 2008: Agriculture for Development," Report, World Bank.

Young, A. (2019): "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results*," *The Quarterly Journal of Economics*, 134, 557–598.

A Additional summary statistics and balance checks

A.1 Balance across various samples

Table A.1: Baseline balance by treatment assignment, primary sample

Variable	(1) Control Mean (SE)	(2) Treatment Mean (SE)	Difference (2)-(1)
Married (0/1)	0.90 (0.01)	0.89 (0.02)	-0.01
Household head has secondary school (0/1)	0.34 (0.02)	0.38 (0.03)	0.01
Household income >50% from farm labor (0/1)	0.79 (0.02)	0.76 (0.03)	-0.02
Used improved ag technology in 2016 (0/1)	0.62 (0.02)	0.66 (0.03)	0.00
Reports knowledge of 1AF practices (0/1)	0.06 (0.01)	0.14 (0.02)	0.07***
Intercropped maize and beans in 2016 (0/1)	0.47 (0.03)	0.56 (0.03)	0.05
Reports having credit access in 2016 (0/1)	0.71 (0.02)	0.73 (0.03)	0.03
Household size	6.63 (0.13)	6.75 (0.15)	0.20
Acres under maize cultivation in 2016	1.00 (0.04)	0.99 (0.05)	-0.01
Maize yield (kg/acre) in 2016	429.55 (21.81)	443.26 (21.75)	-26.34
F-statistic (test of joint significance)			1.25
Number of observations			682

Note: field office area fixed effects included in all balance test regressions. Standard errors are clustered at the treatment assignment (farmer group cluster) level.

Table A.2: Baseline balance by treatment assignment, pre-exposed sample

Variable	(1) Control Mean (SE)	(2) Treatment Mean (SE)	Difference (2)-(1)
Married (0/1)	0.86 (0.01)	0.88 (0.01)	0.01
Household head has secondary school (0/1)	0.41 (0.02)	0.47 (0.02)	0.06*
Household income >50% from farm labor (0/1)	0.77 (0.02)	0.79 (0.02)	0.01
Used improved ag technology in 2016 (0/1)	0.88 (0.01)	0.88 (0.01)	0.00
Reports knowledge of 1AF practices (0/1)	0.71 (0.02)	0.71 (0.02)	0.00
Intercropped maize and beans in 2016 (0/1)	0.48 (0.02)	0.43 (0.02)	-0.02
Reports having credit access in 2016 (0/1)	0.71 (0.02)	0.74 (0.02)	0.01
Household size	6.68 (0.10)	6.98 (0.11)	0.26*
Acres under maize cultivation in 2016	1.01 (0.03)	1.08 (0.04)	0.05
Maize yield (kg/acre) in 2016	602.12 (16.73)	661.56 (21.24)	79.04***
F-statistic (test of joint significance)			1.26
Number of observations			1214

Note: field office area fixed effects included in all balance test regressions. Standard errors are clustered at the treatment assignment (farmer group cluster) level.

A.2 Baseline comparison of pre-exposed and new farmers

Table A.3: Baseline balance across "primary" and pre-exposed samples

Variable	(1) Primary Sample Mean (SE)	(2) Pre-Exposed Sample Mean (SE)	Difference (2)-(1)
Panel A: Pre-Specified Control Variables			
Married (0/1)	0.90 (0.01)	0.87 (0.01)	-0.02
Household head has secondary school (0/1)	0.36 (0.02)	0.44 (0.01)	0.07**
Household income >50% from farm labor (0/1)	0.78 (0.02)	0.78 (0.01)	0.04*
Used improved ag technology in 2016 (0/1)	0.64 (0.02)	0.88 (0.01)	0.21***
Reports knowledge of 1AF practices (0/1)	0.10 (0.01)	0.71 (0.01)	0.59***
Intercropped maize and beans in 2016 (0/1)	0.51 (0.02)	0.46 (0.01)	-0.08***
Reports having credit access in 2016 (0/1)	0.72 (0.02)	0.73 (0.01)	0.06***
Household size	6.68 (0.10)	6.82 (0.07)	0.29**
Acres under maize cultivation in 2016	0.99 (0.03)	1.05 (0.02)	0.09**
Maize yield (kg/acre) in 2016	435.32 (15.59)	629.64 (13.34)	171.79***
Panel B: Additional Baseline Variables			
Female Respondent (0/1)	0.57 (0.02)	0.61 (0.01)	0.02
Income from non-farm labor (0/1)	0.50 (0.02)	0.49 (0.01)	-0.02
Income from business (0/1)	0.67 (0.02)	0.64 (0.01)	-0.01
Income from remittances (0/1)	0.43 (0.02)	0.51 (0.01)	0.09***
Income from formal employment (0/1)	0.26 (0.02)	0.26 (0.01)	0.01
Hired farm labor in 2016 (0/1)	0.72 (0.02)	0.75 (0.01)	0.04
DAP used in 2016 (kgs)	22.64 (1.45)	36.15 (1.03)	12.35***
# of non-maize crops cultivated	0.51 (0.03)	0.56 (0.02)	0.05
# extension officer visits, 2016	0.18 (0.04)	0.22 (0.03)	0.07
# of non-1AF farming org. memberships	0.17 (0.02)	0.15 (0.01)	0.02
Asset score	17.05 (0.27)	20.15 (0.20)	2.79***
F-statistic (test of joint significance)			39.19
Number of observations			1896

Note: field office area fixed effects included in all balance test regressions. Standard errors are clustered at the treatment assignment (farmer group cluster) level.

A.2.1 Distributions of nonbinary baseline variables by treatment status

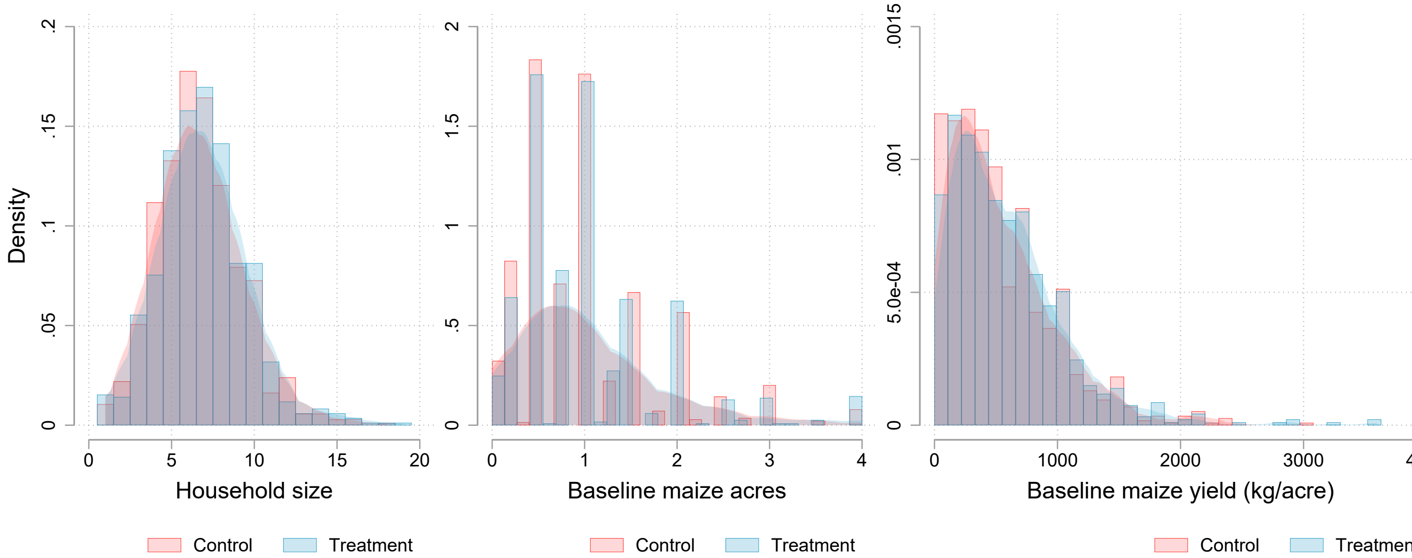


Figure A.1: Distributions of non-binary baseline characteristics, by treatment status

A.3 Baseline 1AF knowledge and practice adoption

Table A.4: Practice adoption and baseline knowledge, primary sample

	Row Spacing		Plant Spacing		Fertilizer Timing	
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline knowledge	0.090	0.121	0.062	0.083	-0.039	-0.046
of 1AF practices	(0.109)	(0.113)	(0.065)	(0.063)	(0.077)	(0.075)
1AF participant		0.241***		0.211***		0.646***
		(0.042)		(0.034)		(0.030)
1AF participant		-0.181		-0.107		0.051
× baseline knowledge of 1AF practices		(0.140)		(0.100)		(0.086)
Observations	395	682	395	682	395	682
R^2	0.061	0.073	0.114	0.142	0.069	0.441
Excl. Group Mean Dep. Var	0.368	0.368	0.086	0.086	0.268	0.268
Sample	Prim. Ctrl.	Primary	Prim. Ctrl.	Primary	Prim. Ctrl.	Primary

Results in this table are from linear regressions of the outcomes (adoption of 1AF-recommended row spacing, plant spacing, and fertilizer timing) on the baseline 1AF knowledge dummy. The excluded group mean is shown for control group farmers with zero baseline knowledge of 1AF practices. Standard errors (in parentheses) clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects.

B Variable construction and measurement

Maize yield measurements were taken by enumerators on two 40-square-meter areas selected from farmer plots. Enumerators marked each area before farmers harvested any maize. For treated farmers, two areas were marked each on enrolled and non-enrolled land. Additionally, if farmers planned to harvest green maize from some parts of their plot, enumerators marked areas within these parts for measurement. Yield variables use the average dry weight from the two marked areas.

Some readers may wish to see what the raw data look like, in addition to the regression results. Figure B.1 shows the distributions of maize yields per acre and total maize output. In the top figure, for treatment farmers we show yields on the portion of the land that farmers enrolled in the 1AF program. In the second figure, we show yields on the remaining land which was not enrolled in the program. The bottom figure sums all maize produced by farmers to show the distribution of total maize output among farmers in each group.

We show effects separately for the two components of profits, revenue and costs, in Table B.1. We calculate projected revenues using average market prices from nearby vendors collected in September 2017. Following the PAP, we multiply prices by 1.08 to account for typical price increases over the consumption/selling season. This is a conservative assumption compared to findings in Burke et al. (2019) who document maize price increases between 42% and 125% in post-harvest seasons from 2013 to 2017 in nearby areas of western Kenya. We also do not ascribe any value to increased storage (due to 1AF encouragement to adopt hermetic PICS bags) or differentially higher prices due to encouragement to store more maize to sell later.

We calculate farmer input costs using program costs and self-reported costs for treated farmers, and self-reported seed and fertilizer costs for control farmers. Labor costs include land prep, plowing, and planting costs, collected in a survey after planting, as well as post-planting costs collected at harvest time. For paid labor, we use farmer self-reported costs by planting phase. To include the opportunity cost of unpaid labor use, we calculate the mean day wage reported within the sample, devalue this mean wage by 50% (roughly the rural unemployment rate according to DHS data), and multiply this devalued mean by total person-days of unpaid labor for each planting phase. We additionally account for land rental costs and program enrollment costs, which include the cost of input delivery and input insurance. Profit is simply the difference between projected farmer revenues and costs.

Tables B.2 and B.3 measure fertilizer use in kilograms rather than valuing their use in USD. We detect a sizeable increase in fertilizer use when measured in kilograms, reinforcing

Table B.1: Total revenue and costs

	Total Revenue		Total Costs	
	(1)	(2)	(3)	(4)
1AF participant=1	100.951***	105.454***	41.281***	42.564***
	(33.194)	(35.180)	(5.456)	(5.689)
Past 1AF participant=1		76.171**		14.472***
		(29.381)		(4.624)
Past 1AF participant=1 × 1AF participant=1		-15.669		4.889
		(47.194)		(7.816)
Observations	682	1896	682	1896
R^2	0.148	0.096	0.230	0.189
Control Mean Dep. Var	425.521	425.521	93.407	93.407

Results in this table are from linear regressions of the outcomes on the treatment dummy. Standard errors (in parentheses) clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects.

that it is unlikely 1AF prices that are driving increased expenditures. We can also break down fertilizer use by phase, and here we see the underlying substitution behind the effect on fertilizer timing noted in Table 3. Farmers in the treatment group are not only using more fertilizer at the "correct" time, but also using less fertilizer at incorrect times.

Table B.2: Quantity of fertilizer used (kgs) by planting phase, primary sample

	At Planting		Post Planting	
	(1)	(2)	(3)	(4)
	DAP	CAN	DAP	CAN
1AF participant	22.250***	-0.054	-7.180***	15.347***
	(1.906)	(0.145)	(1.845)	(2.365)
Observations	682	682	682	682
R^2	0.236	0.013	0.221	0.209
Control Mean Dep. Var	8.976	0.213	10.256	18.119

Results in this table are from linear regressions of the outcomes in each column on the treatment dummy. Standard errors (in parentheses) clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects.

Table B.3: Quantity of fertilizer used (kgs) by planting phase, full sample

	At Planting		Post Planting	
	(1)	(2)	(3)	(4)
	DAP	CAN	DAP	CAN
1AF participant=1	22.201***	-0.035	-7.197***	15.661***
	(1.995)	(0.177)	(1.877)	(2.310)
Past 1AF participant=1	7.708***	0.235	-2.753*	5.379**
	(1.650)	(0.261)	(1.530)	(2.060)
1AF participant=1 × Past 1AF participant=1	-2.773	0.258	3.502*	0.336
	(2.555)	(0.326)	(1.810)	(3.003)
Observations	1896	1896	1896	1896
R^2	0.177	0.009	0.166	0.185
Control Mean Dep. Var	15.044	0.394	8.598	22.091

Results in this table are from linear regressions of the outcomes in each column on the treatment dummy. Standard errors (in parentheses) clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects.

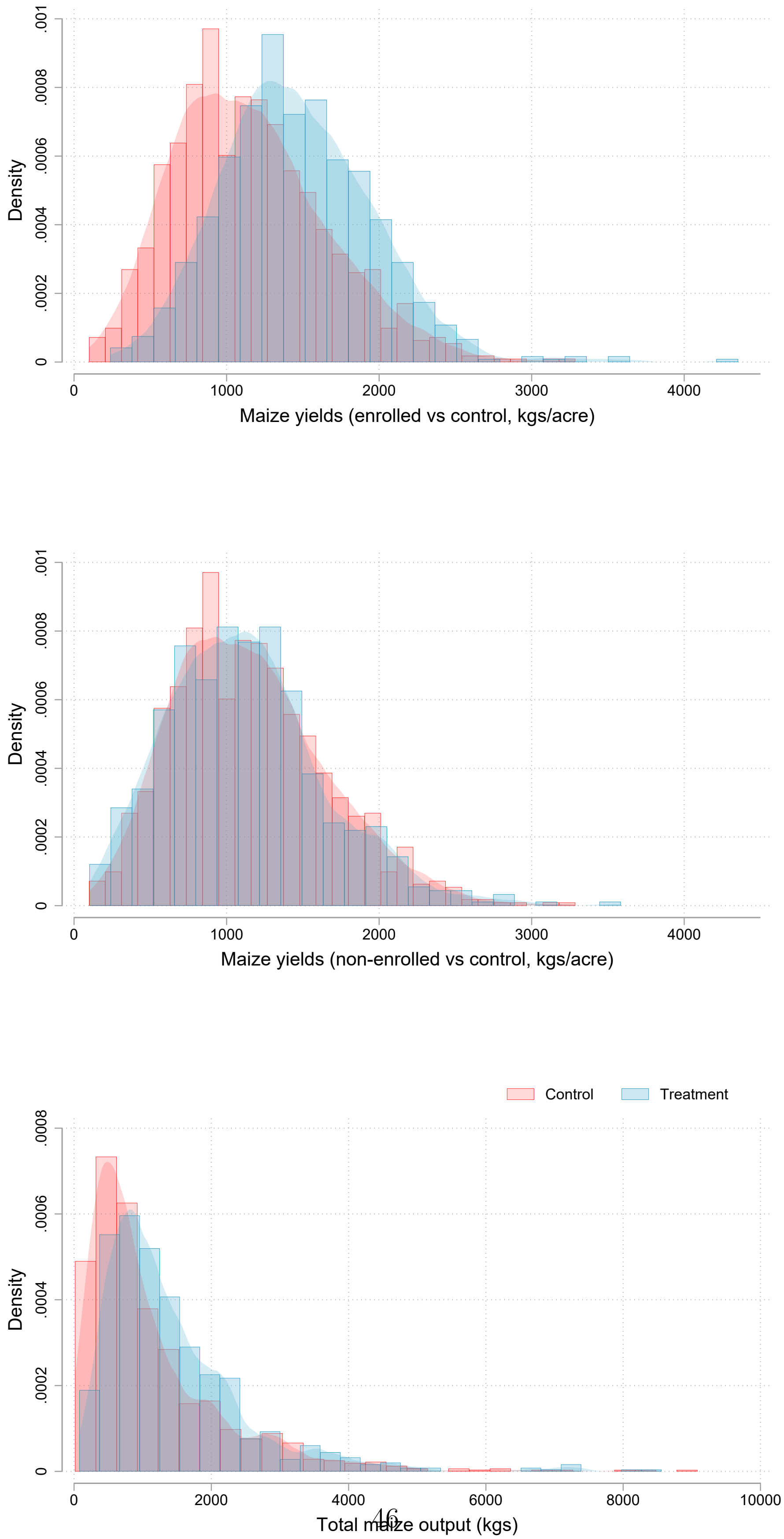


Figure B.1: Distributions of maize yields and output

C Robustness and external validity

C.1 Attrition

This section addresses the different types of attrition that we find in our data: first, we examine attrition that occurred between the baseline survey but before farmers passed the pre-enrollment qualification. To qualify for the study, farmers had to pay a 500 KES deposit and form farmer groups. Farmers who did not complete the prepayment and/or failed to form a group of at least 3 members were dropped. Because the baseline was completed before this qualification stage, the study intentionally sampled more farmers than the desired sample size. Since treatment was assigned after the pre-qualification, we should not expect any threats to internal validity from post-baseline attrition. Understanding what caused this attrition may however be informative with respect to generalizability. We examine this in Section C.1.1.

Conditional on being enrolled in the study, attrition is driven by two main factors: missing land size values and missing harvest data. During data cleaning, we lose 175 observations that are missing land size and 329 observations that are missing dry weight harvest data (of these 329, we observe fresh weight harvest measurements for 53 observations). This attrition is not trivial, and in the following subsections we discuss how we ensure our results are robust to a variety of assumptions about the missing data.

C.1.1 Attrition prior to season and external validity

Table C.1 shows how the qualified and dropped samples differ. The farmers who managed to qualify differ significantly from the farmers who dropped out. Farmers who passed the pre-qualification stage were more likely to be pre-exposed, had better knowledge of 1AF practices, were more likely to use improved seeds and fertilizer, and were more likely to intercrop. Qualified farmers also seem more specialized in farming, as measured by the higher likelihood of receiving more than half their income from farm labor, farming more acres for maize in 2016 and 2015, and the amount of maize that they harvested in previous seasons. Finally, qualified farmers are wealthier, as indicated by ownership of more land and a higher asset score.

These variables suggest that 1AF may not reach the poorest farmers—a common challenge for entrepreneurially-focused agriculture programs (Carter, Tjernström, and Toledo, 2019). That said, compared to a more representative sample—Tegemeo Institute's panel survey of maize farming households—the qualified farmers cultivate fewer acres than the

Table C.1: Baseline balance across enrolled and dropped groups

Variable	(1) Qualified Mean (SE)	(2) Dropped Mean (SE)	Difference (2)-(1)
Panel A: Pre-Specified Control Variables			
Married (0/1)	0.88 (0.01)	0.88 (0.01)	0.00
Household head has secondary school (0/1)	0.40 (0.01)	0.32 (0.02)	-0.08***
Household income >50% from farm labor (0/1)	0.77 (0.01)	0.72 (0.02)	-0.05**
Used improved ag technology in 2016 (0/1)	0.77 (0.01)	0.68 (0.02)	-0.09***
Reports knowledge of 1AF practices (0/1)	0.49 (0.01)	0.34 (0.02)	-0.15***
Intercropped maize and beans in 2016 (0/1)	0.47 (0.01)	0.43 (0.02)	-0.04*
Reports having credit access in 2016 (0/1)	0.71 (0.01)	0.70 (0.02)	0.00
Household size	6.72 (0.05)	6.09 (0.10)	-0.63***
Acres under maize cultivation in 2016	1.01 (0.02)	0.84 (0.03)	-0.17***
Maize yield (kg/acre) in 2016	555.75 (9.17)	470.22 (17.01)	-85.52***
Panel B: Additional Baseline Variables			
Female Respondent (0/1)	0.59 (0.01)	0.51 (0.02)	-0.08***
Income from non-farm labor (0/1)	0.48 (0.01)	0.57 (0.02)	0.10***
Income from business (0/1)	0.64 (0.01)	0.58 (0.02)	-0.06***
Income from remittances (0/1)	0.48 (0.01)	0.46 (0.02)	-0.02
Income from formal employment (0/1)	0.26 (0.01)	0.19 (0.02)	-0.08***
Hired farm labor in 2016 (0/1)	0.72 (0.01)	0.67 (0.02)	-0.05**
DAP used in 2016 (kgs)	30.83 (0.75)	22.69 (1.92)	-8.14***
# of non-maize crops cultivated	0.54 (0.01)	0.46 (0.03)	-0.07**
# extension officer visits, 2016	0.19 (0.02)	0.13 (0.03)	-0.06
# of non-1AF farming org. memberships	0.15 (0.01)	0.10 (0.01)	-0.05**
Asset score	18.75 (0.15)	16.18 (0.26)	-2.56***
F-statistic (test of joint significance)			94.36
Number of observations			3002

Tegemeo sample (see Figure C.1).[20] Note that the Tegemeo survey relies on self-reported acreage; we therefore report our baseline self-reported acreage variable (dotted line in Figure C.1) in addition to the GPS-measured land sized (short-dashed line in Figure C.1). Our sample farmers farm smaller plots than the average TAPRA survey participant.

Our pre-exposed farmers could be driving these differences. To check this, Table C.2 repeats the balance test for the "primary" sample (i.e., without the pre-exposed farmers). Farmers who make it past the pre-qualification stage are still significantly different, but mostly related to wealth and agricultural specialization. In terms of knowledge of 1AF prac-

[20]Tegemeo Institute's panel survey was designed to be broadly representative of the maize-growing regions of Kenya.

Table C.2: Baseline balance across enrolled and dropped groups among primary sample

Variable	(1) Qualified Mean (SE)	(2) Dropped Mean (SE)	Difference (2)-(1)
Panel A: Pre-Specified Control Variables			
Married (0/1)	0.88 (0.01)	0.87 (0.02)	-0.01
Household head has secondary school (0/1)	0.35 (0.02)	0.27 (0.03)	-0.08***
Household income >50% from farm labor (0/1)	0.76 (0.01)	0.68 (0.03)	-0.08***
Used improved ag technology in 2016 (0/1)	0.59 (0.02)	0.57 (0.03)	-0.02
Reports knowledge of 1AF practices (0/1)	0.10 (0.01)	0.10 (0.02)	0.01
Intercropped maize and beans in 2016 (0/1)	0.51 (0.02)	0.47 (0.03)	-0.04
Reports having credit access in 2016 (0/1)	0.67 (0.02)	0.68 (0.03)	0.01
Household size	6.59 (0.09)	5.84 (0.13)	-0.76***
Acres under maize cultivation in 2016	0.95 (0.03)	0.76 (0.04)	-0.19***
Maize yield (kg/acre) in 2016	433.24 (13.78)	398.07 (22.38)	-35.17
Panel B: Additional Baseline Variables			
Female Respondent (0/1)	0.57 (0.02)	0.52 (0.03)	-0.05
Income from non-farm labor (0/1)	0.46 (0.02)	0.56 (0.03)	0.10***
Income from business (0/1)	0.63 (0.02)	0.58 (0.03)	-0.05*
Income from remittances (0/1)	0.42 (0.02)	0.47 (0.03)	0.05
Income from formal employment (0/1)	0.25 (0.01)	0.18 (0.02)	-0.07***
Hired farm labor in 2016 (0/1)	0.68 (0.02)	0.67 (0.03)	-0.01
DAP used in 2016 (kgs)	21.92 (1.28)	14.82 (1.75)	-7.10***
# of non-maize crops cultivated	0.51 (0.02)	0.44 (0.04)	-0.06
# extension officer visits, 2016	0.16 (0.03)	0.14 (0.05)	-0.02
# of non-1AF farming org. memberships	0.15 (0.02)	0.09 (0.02)	-0.06**
Asset score	16.31 (0.26)	14.79 (0.35)	-1.51***
F-statistic (test of joint significance)			46.52
Number of observations			1155

tices, use of improved seed and fertilizer, credit access, and intercropping, the two groups are statistically indistinguishable. We see also that many differences, while significant, are much smaller in magnitude.

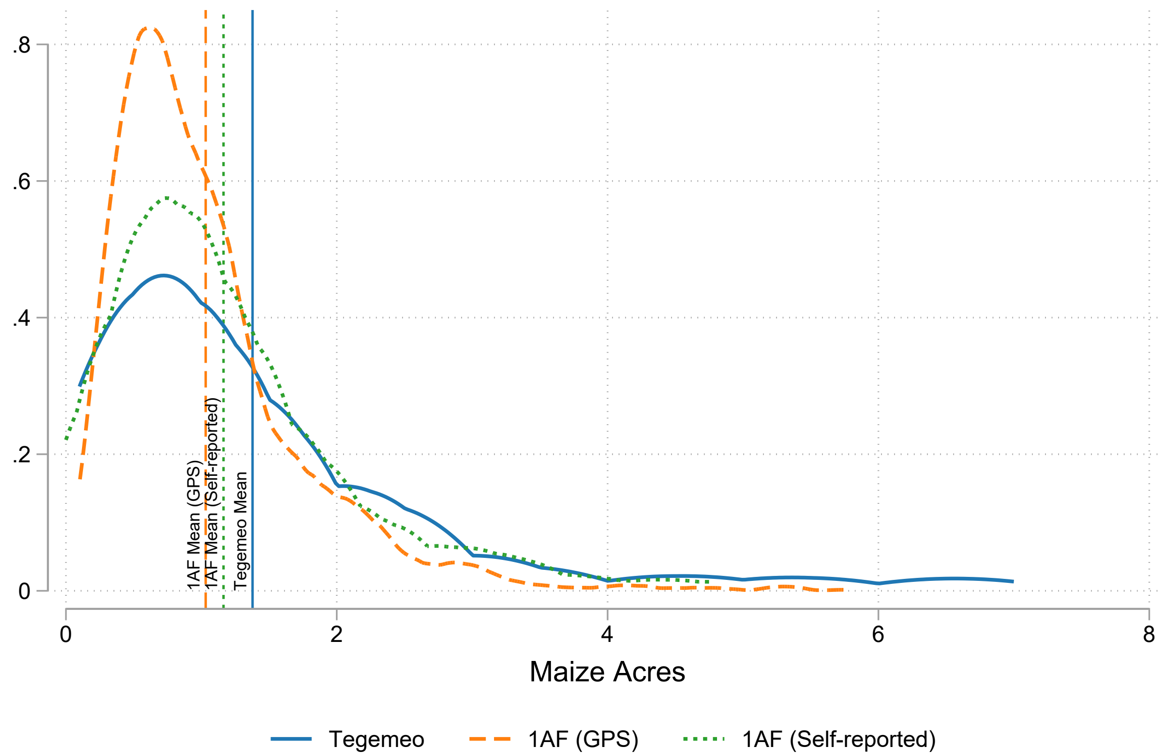


Figure C.1: Comparison of maize acres across study sample and TAPRA survey

C.1.2 Within-season attrition

To test the robustness of our results to the potentially non-random missing data, we impute missing values in a number of ways. The main source of missingness occurs in the harvest variable (yields on the crop-cut plot) and the land size variable. Figures C.2 and C.3 show the results of testing these various strategies for dealing with attrition in our two key variables and compares the results to our main specifications.

For yields, we test two different imputation strategies to compare against the results in our main sample. First, for a subset of farmers, we have non-missing "fresh weight" yields (but missing dry weight yields). We use the sample of farmers for whom we have both fresh and dry maize weights to estimate the relationship between the two variables. We then predict the dry weight yields for the subsample with only fresh weights. For farmers without any harvest weights, we make the most conservative assumption and assume zero yields (i.e., total crop failure). This should allow us to estimate a realistic lower bound on possible treatment effects. Second, we complement the above with a simpler approach where we replace missing yield values with the mean yields by plot type (treated-enrolled plot, treated-non-enrolled plot, control plot).

For missing land size, we test imputing using survey-collected self-reported measures across three different surveys that were part of the data collection efforts. Similarly to the yield imputations, we use the estimated relationships in the sample of households with overlapping GPS-measured land size and self-reported data to predict land size for the missing observations. Here, too, we conduct the predictions separately by plot type. Given that we have several different self-reported measures, we report results from four different imputations, using each self-reported value on its own *(i)-(iii)*, and using the mean of these three self-reports *(iv)*. For the latter, if a participant is missing one of the three self-reports, we take the mean of the non-missing self-reported values.

Figure C.2: Treatment effect estimates on maize yields with various methods of accounting for within-season attrition

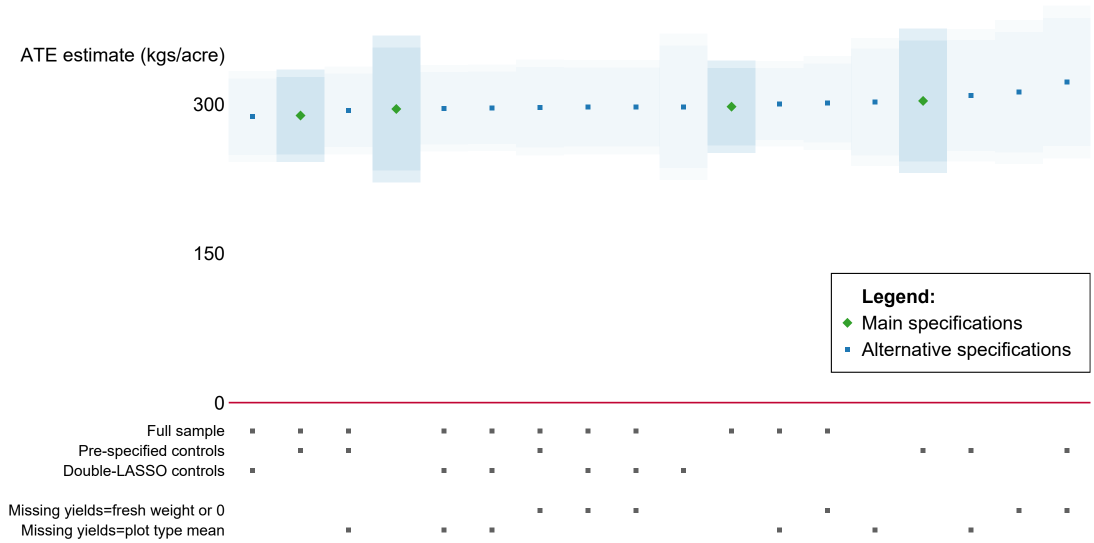
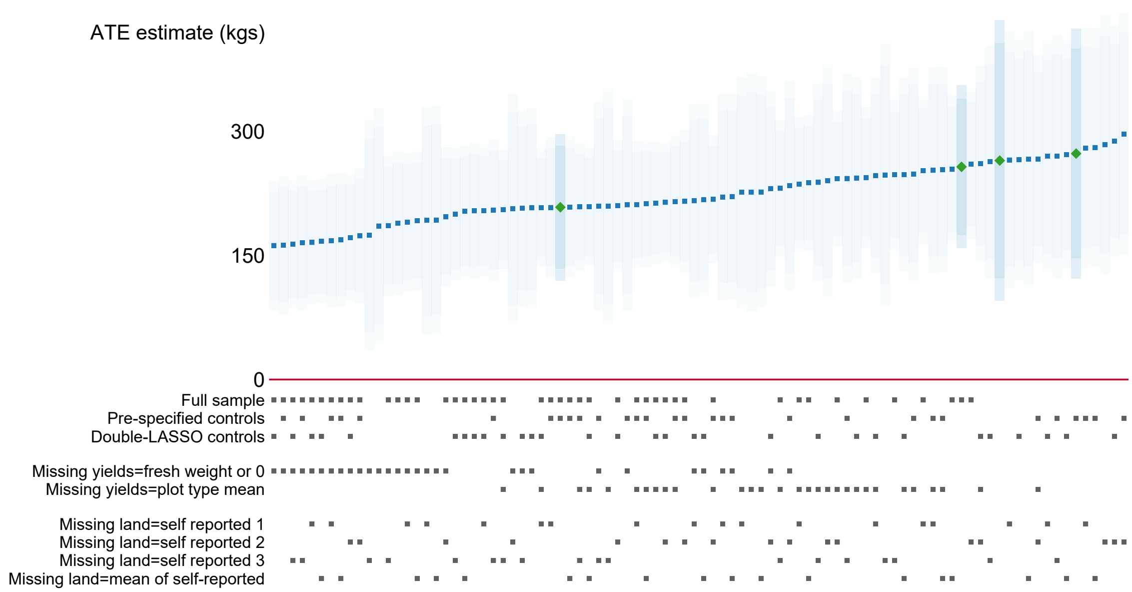


Figure C.3: Treatment effect estimates on total maize output with various methods of accounting for within-season attrition



C.2 Alternative specification details

Table C.3: Yields and output with pre-specified controls

	Maize yields enrolled vs. control		Maize yields non-enrolled vs. control		Total maize output	
	(1)	(2)	(3)	(4)	(5)	(6)
1AF participant	303.99***	308.68***	-51.18	-42.24	274.15***	270.66***
	(36.810)	(36.910)	(35.680)	(39.250)	(75.820)	(81.870)
Pre-exposed		-9.72		-29.31		70.48
		(37.800)		(37.510)		(71.780)
1AF participant × pre-exposed		-29.67		-20.64		-93.74
		(49.940)		(54.990)		(113.330)
1AF + (*1AF* × *Pre-exposed*)		279.01***		-62.88		176.92***
		(31.31)		(41.25)		(62.34)
Control group mean	1128.39	1150.49	1128.39	1150.49	1082.38	1164.00
Observations	682	1896	614	1701	682	1896

Note: This table presents results from linear regressions of the outcomes at the top of each column on the treatment dummy. Maize yields are measured as kgs of maize per acre cultivated, and total maize output is measured in kgs. In columns 1 and 2, enrolled refers to the land farmers in the treatment group enrolled in the 1AF program, which may be less than the full amount of land on which they cultivate maize. In columns 3 and 4, we include only farmers who also cultivated maize on land they did not enroll with 1AF and compare the yields from that non-enrolled land to yields from control group farmers. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. Columns 1, 3, and 5 include only farmers who had never previously participated in the 1AF program. Columns 2, 4, and 6 additionally include a sample of farmers who had previously enrolled in the 1AF program (see Section 2.2). All regressions include strata (field office) fixed effects. This table replicates Table 2 above but includes the set of pre-specified controls. Binary baseline controls are married, at least secondary education by male household head, whether the household receives half of its income from farm labor, whether the household used improved seeds at baseline, whether the household reported knowledge of 1AF practices at baseline, whether the farmer intercropped beans and maize at baseline, and whether the household reported having access to credit. Continuous baseline controls are household size, a measure of Fall Army Worm presence collected during the 2017 season, baseline maize acres, and baseline maize yields per acre.

Although the PAP pre-specified a set of baseline controls that should be included in analysis of our outcomes of interest, the baseline survey contains a number of additional covariates that may be useful in improving the precision of treatment effect estimates.[21] Following Belloni et al. (2014) and Urminsky et al. (2016), we implement a double-Lasso

[21] An additional deviation from the original PAP is that the PAP included an additional covariate measuring Fall Army Worm incidence. During the course of the 2017 season, a pest called Fall Army Worm had a dramatic effect on Kenyan farmers. We were concerned that the pest may affect treatment estimates. To measure the extent of FAW, enumerators visited farmer fields during the growing season and randomly selected 30 plants to inspect and check for signs of Fall Army Worm. The problem was indeed widespread: nearly 80% of farmers had at least some plants affected by FAW, and 66% of farmers had signs of FAW on all inspected plants. In theory, we may be concerned that FAW incidence could be higher among treated farmers, for example if their maize is healthier and more easily allows FAW to propagate. In practice, however, including this variable as an additional control has no meaningful impact on the magnitude or precision of our treatment effect estimates.

Table C.4: Profits with pre-specified controls

	Profit (PAP Definition)		Profit (Full Labor Costs)	
	(1)	(2)	(3)	(4)
1AF participant	63.56**	60.73**	59.88**	57.25*
	(27.530)	(30.220)	(27.490)	(30.030)
Pre-exposed		29.11		30.32
		(26.600)		(26.280)
1AF participant × pre-exposed		-39.76		-41.36
		(41.170)		(40.900)
1AF + (*1AF* × *Pre-exposed*)		20.96		15.89
		(22.97)		(22.83)
Control group mean	332.11	361.58	314.29	344.92
Observations	682	1896	682	1896

Note: This table presents results from linear regressions of the outcomes shown at the top of each column on the treatment dummy. Profits are calculated as the value of maize production less costs (see Appendix B), and are reported in USD. Columns 1 and 2 follow the PAP and account for unpaid labor costs by valuing unpaid time at 50% of the average day-wage reported in the sample for paid labor. Columns 3 and 4 instead value unpaid labor at the full average day-wage reported in the sample. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. Odd columns include only farmers who had never previously participated in the 1AF program. Even columns additionally include a sample of farmers who had previously enrolled in the 1AF program (see Section 2.2). All regressions include strata (field office) fixed effects. This table replicates Table 6 above but includes the set of pre-specified controls. Binary baseline controls are married, at least secondary education by male household head, whether the household receives half of its income from farm labor, whether the household used improved seeds at baseline, whether the household reported knowledge of 1AF practices at baseline, whether the farmer intercropped beans and maize at baseline, and whether the household reported having access to credit. Continuous baseline controls are household size, a measure of Fall Army Worm presence collected during the 2017 season, baseline maize acres, and baseline maize yields per acre.

procedure for variable selection. In addition to all pre-specified controls and fixed effects, we additionally include respondent gender, DAP use, crop diversity, extension officer visits, membership in other farming organizations, use of hired farm labor, an assets score, and dummies for income from nonfarm labor, businesses, remittances, or formal employment. Variable selection is distinct for each outcome variable and each sample definition. The results of this exercise, presented in Table C.5, demonstrate that the results are not particularly sensitive to the inclusion of an optimal set of controls.

Table C.5: Yields and output with double-Lasso-selected controls

	Maize yields enrolled vs. control		Maize yields non-enrolled vs. control		Total maize output	
	(1)	(2)	(3)	(4)	(5)	(6)
1AF participant	299.67***	306.46***	-53.32	-37.69	266.08***	253.09***
	(37.190)	(36.580)	(38.080)	(40.260)	(73.720)	(77.590)
Pre-exposed		-7.57		-20.40		62.58
		(37.820)		(37.430)		(69.960)
1AF participant × pre-exposed		-24.46		-24.68		-72.86
		(49.660)		(56.590)		(109.440)
$1AF + (1AF \times \mathit{Pre\text{-}exposed})$		282.00***		-62.38		180.23***
		(31.15)		(40.73)		(59.68)
Control group mean	1128.39	1150.49	1128.39	1150.49	1082.38	1164.00
Observations	682	1896	614	1701	682	1896

Note: This table presents results from linear regressions of the outcomes at the top of each column on the treatment dummy. Maize yields are measured as kgs of maize per acre cultivated, and total maize output is measured in kgs. In columns 1 and 2, enrolled refers to the land farmers in the treatment group enrolled in the 1AF program, which may be less than the full amount of land on which they cultivate maize. In columns 3 and 4, we include only farmers who also cultivated maize on land they did not enroll with 1AF and compare the yields from that non-enrolled land to yields from control group farmers. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. Columns 1, 3, and 5 include only farmers who had never previously participated in the 1AF program. Columns 2, 4, and 6 additionally include a sample of farmers who had previously enrolled in the 1AF program (see Section 2.2). All regressions include strata (field office) fixed effects. This table replicates Table 2 above but includes controls selected with double-Lasso-selected controls (Belloni et al., 2014). The Lasso regressions select from all baseline covariates included in Table C.3, as well as the following additional baseline covariates: respondent gender, DAP use, crop diversity, extension officer visits, membership in other farming organizations, use of hired farm labor, an assets score, and dummies for income from nonfarm labor, businesses, remittances, or formal employment. Variable selection is distinct for each outcome variable and each sample definition.

Table C.6: Profits with double-Lasso-selected controls

	Profit (PAP Definition)		Profit (Full Labor Costs)	
	(1)	(2)	(3)	(4)
1AF participant	61.29**	54.66*	57.42**	50.99*
	(27.080)	(28.960)	(26.980)	(28.770)
Pre-exposed		26.03		27.21
		(26.080)		(25.760)
1AF participant		-31.73		-33.15
× pre-exposed		(39.940)		(39.640)
1AF + (*1AF* × *Pre-exposed*)		22.92		17.84
		(22.15)		(22.01)
Control group mean	332.11	361.58	314.29	344.92
Observations	682	1896	682	1896

Note: This table presents results from linear regressions of the outcomes shown at the top of each column on the treatment dummy. Profits are calculated as the value of maize production less costs (see Appendix B), and are reported in USD. Columns 1 and 2 follow the PAP and account for unpaid labor costs by valuing unpaid time at 50% of the average day-wage reported in the sample for paid labor. Columns 3 and 4 instead value unpaid labor at the full average day-wage reported in the sample. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. Odd columns include only farmers who had never previously participated in the 1AF program. Even columns additionally include a sample of farmers who had previously enrolled in the 1AF program (see Section 2.2). All regressions include strata (field office) fixed effects. The Lasso regressions select from all baseline covariates included in Table C.4, as well as the following additional baseline covariates: respondent gender, DAP use, crop diversity, extension officer visits, membership in other farming organizations, use of hired farm labor, an assets score, and dummies for income from nonfarm labor, businesses, remittances, or formal employment. Variable selection is distinct for each outcome variable and each sample definition.

C.3 Spillovers to non-enrolled land

Table C.7: Behavioral changes, non-enrolled vs control

	Row Spacing		Plant Spacing		Fertilizer Timing	
	(1)	(2)	(3)	(4)	(5)	(6)
1AF participant	0.09*	0.09**	0.03	0.03	0.11***	0.10**
	(0.040)	(0.040)	(0.030)	(0.030)	(0.040)	(0.040)
Pre-exposed		0.07**		0.02		0.13***
		(0.030)		(0.020)		(0.040)
1AF participant × pre-exposed		-0.09		-0.04		-0.11*
		(0.060)		(0.040)		(0.050)
1AF + (*1AF* × *Pre-exposed*)		-0.01		-0.01		0.00
		(0.03)		(0.02)		(0.03)
Control group mean	0.37	0.41	0.09	0.13	0.27	0.38
Observations	614	1701	614	1701	614	1701

Note: This table presents results from linear regressions of the outcome variables in each column on the treatment dummy. For treated farmers, we consider practices only on their non-enrolled land (if any). For control group farmers, we consider practices on all of their land. Row and plant spacing are indicator variables for whether or not a farmer was within 5cm of the recommended spacing, for the rows in which they planted and the plants within rows, respectively. Fertilizer timing is an indicator variable for whether or not farmers applied DAP at planting and CAN at top dressing (the timing recommended by the NGO). Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. Columns 1, 3, and 5 include only farmers who had never previously participated in the 1AF program (the "primary" sample). Columns 2, 4, and 6 additionally include a sample of farmers who had previously enrolled in the 1AF program (the "full" sample; see Section 2.2). All regressions include strata (field office) fixed effects.

Table C.8: Behavioral changes, enrolled vs. non-enrolled, treated sample only

	Row Spacing		Plant Spacing		Fertilizer Timing	
	(1)	(2)	(3)	(4)	(5)	(6)
Enrolled plot	0.13**	0.13**	0.17***	0.17***	0.53***	0.53***
	(0.050)	(0.050)	(0.040)	(0.040)	(0.040)	(0.040)
Pre-exposed		-0.02		-0.02		0.02
		(0.050)		(0.030)		(0.040)
Enrolled plot		0.02		-0.03		-0.05
× pre-exposed		(0.060)		(0.040)		(0.050)
Enrolled + (*Enrolled* × *Pre-exposed*)		0.15***		0.14***		0.48***
		(0.03)		(0.02)		(0.03)
Non-enrolled plot group mean	0.43	0.42	0.15	0.14	0.39	0.41
Observations	438	1308	438	1308	438	1308

Note: This table presents results from linear regressions of the outcome variables in each column on the enrolled plot dummy. The sample includes only treatment group farmers who farmed more maize land than they enrolled, and compares practices on enrolled and non-enrolled land. Row and plant spacing are indicator variables for whether or not a farmer was within 5cm of the recommended spacing, for the rows in which they planted and the plants within rows, respectively. Fertilizer timing is an indicator variable for whether or not farmers applied DAP at planting and CAN at top dressing (the timing recommended by the NGO). Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. Columns 1, 3, and 5 include only farmers who had never previously participated in the 1AF program (the "primary" sample). Columns 2, 4, and 6 additionally include a sample of farmers who had previously enrolled in the 1AF program (the "full" sample; see Section 2.2). All regressions include strata (field office) fixed effects.

D Additional heterogeneity results

Beyond the heterogeneity results found in Section 5.3, another way to view heterogeneity is through distribution regressions (Chernozhukov, Fernández-Val, and Melly, 2013). These are related to unconditional quantile partial effects but can be easier to interpret since the x-axis shows meaningful values rather than percentiles of a distribution. We implement this by running a series of regressions of the following form

$$1(Y > x)_{ij} = \alpha + \beta_x T_i + \gamma_j + \epsilon_i,$$

where Y is the outcome of interest, x varies along the x-axis of each figure along the support of the outcome variable, T_i is the treatment dummy, and γ_j is a cluster fixed effect. Each blue dot in Figure D.1 is an estimated β_x coefficient. The results shown here are from estimations with a linear probability model, but the results are robust to using a logit model to estimate the threshold probabilities.

The results for maize yields and total output both suggest that the effects are largest around the lower end of the distribution and they attenuate at large values of the outcomes. Maize yields likely have some physiological upper bound, beyond which decreasing marginal returns to inputs start to make additional intensification less effective absent further investment in mechanization or land consolidation.

A key difference between the Chernozhukov et al. (2018) approach and the distribution regression approach of Chernozhukov et al. (2013) is that the former focuses on whether specific covariates can predict the size of participants' treatment effects. Figure 2 presents the Sorted Group Average Treatment Effects (GATES) estimated using Neural Nets. Each vertical bar represents the estimated treatment effect at a different percentile of the predicted treatment effect distribution.

For the groups who are predicted to have low treatment effects based on observables, the GATES estimate is only significantly greater than zero for program maize yields. That said, the least-affected and most-affected groups do not differ starkly from each other and the method introduces substantial noise in our relatively small sample. While we do not detect much heterogeneity along the distribution of the outcome variable, this does not automatically rule out the existence of subgroups for whom the treatment is more or less effective. Since this approach relies on covariates to predict heterogeneity, it could nevertheless be the case that we are not including the correct covariates. We test the method with both our pre-specified set of controls and a larger set of covariates which we also use for our double-Lasso methods discussed above in Appendix C.2. In both cases we fail to reject the null hypothesis that the heterogeneity loading parameter differs from zero.

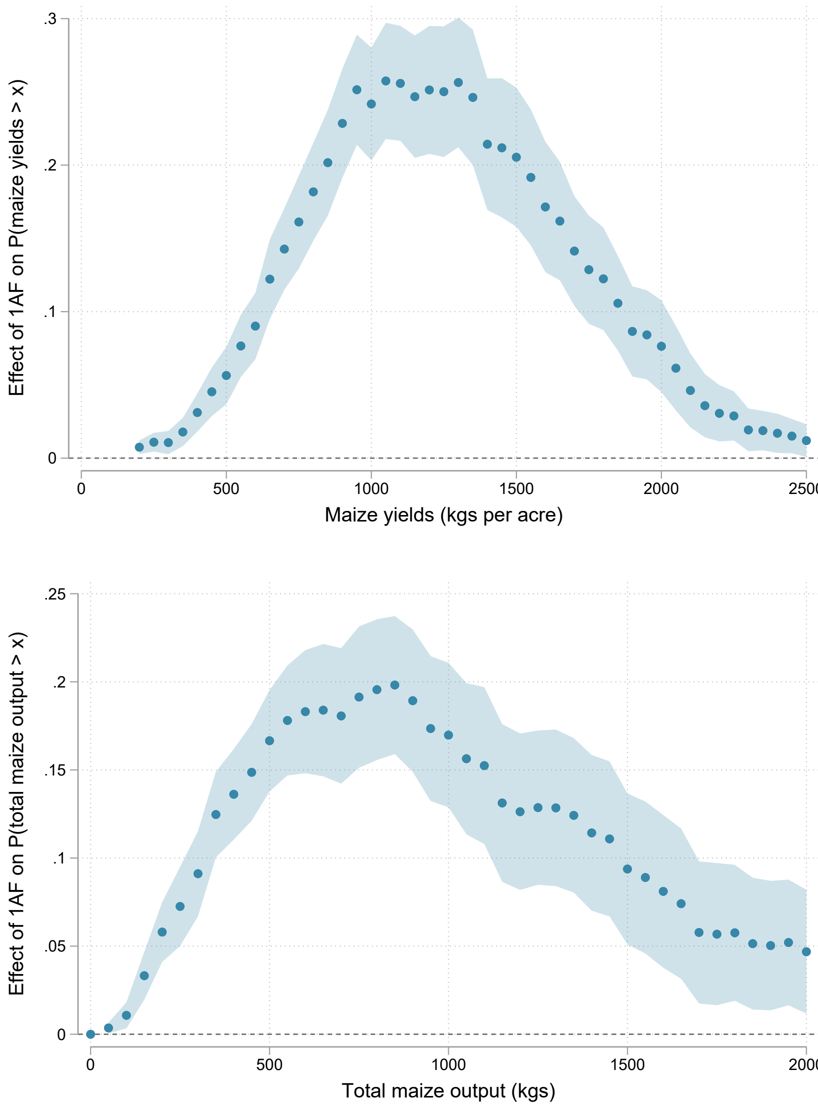


Figure D.1: Distribution regression results

E Participation decisions a year after the experiment

We match our sample farmers to 1AF administrative data and examine their choices in the subsequent season. The goal of this analysis is to shed some light on the mechanisms through which the program works, recognizing that these results are only suggestive.

Table E.1: Enrollment in the 1AF program in the subsequent season

	Enrolled 2018		Acres enrolled 2018	
	(1)	(2)	(3)	(4)
1AF participant	0.00	0.01	0.15***	0.13***
	(0.03)	(0.03)	(0.04)	(0.04)
Pre-exposed		0.07**		0.03
		(0.03)		(0.04)
1AF participant × pre-exposed		-0.02		0.02
		(0.04)		(0.05)
Observations	682	1896	441	1267
R^2	0.332	0.438	0.140	0.101
Control Mean Dep. Var	0.66	0.66	0.62	0.62

Results in this table are from linear regressions of the outcome variables on the treatment dummy. The outcome variable in columns 1 and 2 is a dummy equal to one if the farmer enrolled in the subsequent (2018) season following the experiment, and the outcome in columns 3 and 4 is the enrolled land size recorded in the 2018 administrative data. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects.

First, we ask whether treated farmers are more or less likely to enroll in the subsequent season than the control farmers. We do not have strong *a priori* expectations about the extensive margin effect. If treated farmers learn that program resources are more valuable than they thought, they may be more likely to re-enroll than control farmers. If, instead, the program helps farmers "graduate" by nudging them across a poverty trap threshold, then they might be able to save their increased earnings from the program and hence not need the program in the future. We would see this in the data as treated farmers being less likely to re-enroll the following year. Control group farmers also received a small discount to enroll in the 2018 season as compensation for being randomized out of the program. This could result in a higher enrollment rate for control farmers.

Columns 1 and 2 of Table E.1 show that being randomly allocated to participation in 2017 did not significantly affect the probability that farmers enrolled in 2018, neither in the primary sample (column 1) nor in the pre-exposed sample (column 2). Across the two samples, the 2018 enrollment rate is consistent at around 66 percent.

In addition to the extensive-margin decision, farmers choose how much land to enroll. Larger enrollments include more credit and larger quantities of inputs, but require farmers to commit more land to a specific crop and a larger participation fee. Columns 3 and 4 of Table E.1 show that, conditional on enrolling, farmers in the treatment group increase the amount of land that they enroll in the subsequent year.

These findings are consistent with several mechanisms. Participation in the 1AF program may relax participants' credit constraints, or it may help farmers learn about the returns to program participation and agricultural intensification. We find a similar pattern among pre-exposed farmers with one and two years of past program experience (results available upon request). Although pre-exposure is non-random, the exposure to an additional year is randomly assigned in the pre-exposed sample. Among these repeat-participants, we see increased enrollments which decline based on the number of years of past participation.

F Audit procedures

To boost the credibility of the data collection, 1AF contracted Intermedia Development Consultants (iDC), an independent survey firm, to carry out a three-step audit of the data collection. The full iDC audit report is available from the authors upon request; the overall conclusion of the audit stated the following "With respect to the Teso trial, the strategy and planning are appropriate to the situation, and they have attempted to make every effort to obtain accurate, reliable and valid results. ... Overall, the data collection is well planned and executed. The possibilities for improving performance are quite limited. The survey team members are well selected and committed to the process. The recording and transmission of data is well done, with minimal errors."

The goal of the audit exercise was to evaluate 1AF against best-practice M&E standards, on dimensions including the technical capacity of monitoring staff, supervision, the M&E strategy and planning, data collection, data quality (accuracy, reliability and validity), data recording and analysis, dissemination and use of data, and participation of stakeholders. The auditors recorded observations in the field, which were then compared with data collected by the enumerators to expose any gaps or strengths in the OAF system.

The audit procedure took place in three steps: Step 1 reviewed the planting compliance and crop mix survey data collection. The audit covered a sample of the 2,425 farmers surveyed. Step 2 covered data collection of the beans harvest survey, covering a sample of around 1,400 farmers. Step 3 entailed completing similar activities for the maize harvest survey, covering a sample of around 2,425 farmers.

As part of this process, the audit team participated in group meetings held by the 1AF monitoring team and spent 26 days in the field, during which they observed the work of enumerators on 76 occasions and enumerators' interactions with farmers on 246 occasions. The audit team observed the work of supervisors on 11 separate occasions and that of supervisors as back-checkers on five occasions (Intermedia Development Consultants, 2017). They further carried out parallel data collection efforts to the 1AF data collection, comparing the results and finding minimal discrepancies.

Table F.2: Materials reviewed during iDC audit

Step 1	Step 2	Step 3
Back-check code book	Questionnaires: audited beans dry weight survey	Questionnaire: Harvest survey back-check
Planting compliance and crop mix code book	Questionnaires: for audited bean fresh weight survey	Maize box check
Back-check strategy	Questionnaires: audited harvest box survey	Fresh maize harvest questionnaire
Data collection process and flow	Teso Trial baseline report	Dry weight harvest questionnaire
Teso Trial sites	Teso Trial design and analysis plan	Supervisors checklist
Updated guidance on back-checks	feedback Receipt for beans harvest	List of enumerators and their contacts
Kenya feedback receipts		List of back-checkers and their contacts
Teso Trial baseline report		List of supervisors and contacts
Analysis and design plan		Teso Trial planting compliance crop mix back-check analysis
Audited cases April 4-15		