Can smallholder extension transform African agriculture?

Maya Duru ⓡ Kim Siegal ⓡ
Emilia Tjernström ⓡ Joshua W. Deutschmann *

February 25, 2021

Abstract

Agricultural productivity in Sub-Saharan Africa lags behind all other regions of the world. Decades of investment in agricultural research and extension have yielded more evidence on what fails than on what works—especially for the small-scale producers who dominate the sector. We study a program that targets multiple constraints to productivity at once, similar to anti-poverty "graduation" interventions. Analyzing a randomized controlled trial in western Kenya, we find that participation causes statistically and economically significant gains in output, yields, and profits. In our preferred specification, the program increases maize production by 26% and profits by 16%. The program increases yields uniformly across the sample, while treatment effects on total output and profit impacts are slightly attenuated at the top end of the distribution.

*Deutschmann: University of Wisconsin-Madison, jdeutschmann@wisc.edu. Duru: J-PAL, mduru@povertyactionlab.org. Siegal: One Acre Fund & George Washington University, kim.siegal@oneacrefund.org. Tjernström: University of Sydney, emilia.tjernstrom@sydney.edu.au. The authors thank Kibrom Abay, Brad Barham, Chris Barrett, Lori Beaman, Leah Bevis, 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 benefitted 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 the UW-Madison Agricultural and Applied Economics development workshop. All remaining errors are our own. Strathmore University's IRB approved the study (IRB Approval Number: SU/IRB 0062/16). We registered the trial with the AEA Social Science Registry (AEARCTR-0006675).

How should scarce public resources be allocated across sectors of the economy to advance growth? 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. While agriculture's role in driving growth is theoretically ambiguous (Matsuyama, 1992), a growing body of empirical evidence links increased staple yields to greater national welfare (Bravo-Ortega and Lederman, 2005; Ligon and Sadoulet, 2011; Ravallion and Chen, 2007) and structural change (McArthur and McCord, 2017). More recently, Gollin, Hansen and Wingender (2018) show that agricultural productivity shocks—as measured by the adoption of improved crop varieties—generate large per-capita GDP increases.

Investments in the agricultural sector in sub-Saharan Africa (SSA) have yielded considerable evidence on what fails and a limited catalog of successes. Despite substantial government and donor expenditures on research and extension programs, agricultural input subsidies, and insurance products, agricultural productivity in SSA still lags behind all other regions of the world (Block, 2014; World Bank, 2008). The fact that expenditures have failed to transform the agricultural sector in SSA serves as a kind of Rorschach test for researchers and policy-makers alike. To some, widening agricultural productivity gaps indicate that public resources would be better spent elsewhere. To others, the existence of constraints limiting technology adoption and commercialization suggests that relaxing those constraints with well-designed and targeted investments could stimulate dramatic increases in African smallholder agricultural productivity and economic growth (Byerlee, de Janvry and Sadoulet, 2009; Magruder, 2018).

Our paper contributes to this policy debate by documenting that a multi-faceted program can generate economically meaningful returns. Despite the importance of the agricultural sector, there is a dearth of rigorous empirical evidence on agricultural extension service effectiveness. High-quality evaluations of programs operating at scale are especially rare. Research efforts often suffer from issues of measurement, selection, and comparability across programs (Aker, 2011; Anderson and Feder, 2007). Relatively few experimental studies find positive impacts on farmer practices or improved input use—and even then, researchers often fail to detect measurable increases in yields (Udry et al., 2019) or profits (Beaman et al., 2013).

We report experimental results from a cluster-randomized RCT of a program that targets small-scale staple crop producers: the One Acre Fund (1AF) program. 1AF is a non-governmental organization (NGO), established in 2006, that currently works with over one million farm households in seven countries across Eastern and Southern Africa. We study 1AF's programming by randomizing access to their scaled-up operations in Kenya. We find that program participation causes statistically and economically significant increases in output, yields, and profits. In our preferred specification, total maize output increases by 26% and profits increase by 16%. Our estimated treatment effects are robust across empirical specifications.

Our paper builds on the literature studying technology diffusion in low-income countries. Within this body of work, three constraints stand out as key barriers to adoption of improved agricultural technologies: credit, risk, and information (Feder, Just and Zilberman, 1985; Magruder, 2018). The components of 1AF's programming are closely tied to these constraints: participating farmers receive input loans for high-quality seeds and fertilizer, crop insurance, and training on improved farming practices. None of the ingredients of the program are based on radical ideas; 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).

These similarities aside, the 1AF program distinguishes itself from its antecedents by tightly bundling the component parts into a single unified package (Tinsley and Agapitova, 2018). This kind of bundling is not typical of government-led extension, nor is it common in the economics literature on technology adoption. Since bundled interventions make it difficult to disentangle the underlying economic mechanisms, the academic literature tends to evaluate the impact of relaxing a single constraint at a time—this is especially true in the growing experimental literature reviewed by Magruder (2018).

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 (Banerjee et al., 2015; Bandiera et al., 2017; Balboni et al., 2020). Our results echo recent findings from multi-faceted anti-poverty programs, where individual intervention components seem unable to generate the gains achieved by the full program.

One interpretation of our results is that past research may have underestimated the potential for well-designed programs to increase agricultural productivity. To the extent that governments and donors make their investment choices based on evidence, these findings could lead to more efficient decision-making. We discuss some candidate explanations for why our findings differ from those in the literature in Section 5. Disentangling among these and establishing the external validity of our findings would be fruitful avenues for future research.

In light of the recent proliferation of social enterprises offering financial and extension services to farmers, we undertake two sets of complementary analyses. First, we examine whether the average treatment effects mask underlying impact heterogeneity using econometric and machine learning techniques. Treatment effect heterogeneity could indicate potential efficiency gains from programmatic changes or improved participant targeting. Perhaps surprisingly, treatment effects on maize yields are large and consistent across the sample. The treatment effects on total maize output and profits are lower at the top end of the distribution. The relative lack of heterogeneity may be due to the small upfront fee that 1AF requires prior to enrollment—another difference between 1AF and its many predecessors. If the fee screens potential participants and drives positive selection into the program, this could explain the homogeneous impacts.[1]

Second, we match our sample to administrative records from the subsequent growing season to study participant enrollment decisions at the extensive and intensive margins. Being randomly assigned to the program in 2017 does not affect enrollment probabilities in 2018. However, farmers who are randomly exposed to the program in 2017 enroll significantly more land in the program the following year. These effects are consistent with a successive easing of constraints as well as with participants learning about the returns to the program through participation.

1 Context, data, and experimental design

1.1 One Acre Fund's program in Kenya

We analyze the main operating model of an established agricultural NGO, One Acre Fund. 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 "market bundle" provides farmer groups with group-liability loans for improved seeds and fertilizer, regular training on modern agricultural techniques, crop and funeral insurance, and market facilitation support to help farmers obtain higher prices for

[1] Studies of health products in low-income countries generally find cost-sharing mechanisms to be ineffective screening devices (Ashraf, Berry and Shapiro, 2010; Cohen and Dupas, 2010; Tarozzi et al., 2014). In contrast, Beaman et al. (2020) find evidence of positive selection into an agricultural credit program in Mali.

their output (Tinsley and Agapitova, 2018). Farmer groups are organized by geographical areas and comprised of 8-12 farmers.

The program's focus on credit, risk, and information has support in the economics literature, which has accumulated substantial evidence that failures in these domains hinder farmers' ability or willingness to adopt improved agricultural technologies (Feder, Just and Zilberman, 1985; Magruder, 2018). 1AF's 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 (Bold et al., 2017; Tjernström, Carter and Lybbert, 2018).

Prior to enrolling, participants sort into self-selected farmer groups. The credit component begins with each participant choosing how much of their land to enroll; 1AF determines the size of agricultural input loans as a function of the amount of land enrolled. Farmer groups are jointly responsible for repayment of these loans. Loan terms are relatively flexible, allowing repayment at any time during the growing season. Groups must complete repayment in full within a two-week grace period after harvest.[2] 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).

This experiment was conducted 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 et al., 2011).

Although small-scale farmers in Kenya produce a range of crops across a diverse production environment, 1AF's efforts are focused 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

[2] Historically, repayment rates exceed 97% (Tinsley and Agapitova, 2018).

remains a net importer of maize, despite policy targets to the contrary.

1.2 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.[3] Once participants had self-selected into groups of 8-12 farmers, the randomization was conducted at the level of clusters, which consisted of 2-4 of these joint-liability farmer groups. This aggregation was designed to minimize potential spillovers by maximizing the distance between clusters.

After participants had paid the fee and signed the contract, they were informed of the randomization, which took place as a public lottery. The control group farmers, who were randomly assigned to delay program access received a compensation package consisting of a bundle of household goods valued at the cost of participation. The control group also received a discount for enrollment the following season.

Dealing with control group 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 an area where 1AF was already working but where a large population had not yet been offered the program.

In practice, although 1AF had never offered its program to the sampled villages, some farmers 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. We may nonetheless 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 the "primary sample," which refers to the sample of farmers who had never participated in 1AF programming.

[3] The participation criteria for the standard program is possession of a phone number and national identification. Study participants 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.

The sign of any potential bias 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.

We also do not know the sign of the correlation between self-selection into early access and returns to enrollment. If the most eager farmers have higher returns, we would expect the impacts on our primary sample of hold-out farmers to be a lower bound of the true impacts. If these hold-out farmers instead 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.

1.3 Data

The data collection was managed by the NGO. The evaluation team took several steps to reduce concerns about research independence: 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 included back-checks and in-field supervision. Third, 1AF contracted an independent firm, Intermedia Development Consultants, to carry out a three-step audit of the data collection. Appendix A documents 3ie's review of the pre-analysis plan, and Appendix G provides more details about the audit procedures and results.

Fourth, the main dependent variables, yield and land size, were physically weighed and measured with GPS, respectively, which should minimize the potential for social desirability bias.[4] All weighing and surveying was carried out by 1AF enumerator staff from outside the study area who did not know the sample farmers. Finally, two of the authors on this paper were brought in as independent evaluators. This part of the team reviewed the pre-analysis plan (PAP) prior to data collection and independently conducted the data cleaning, variable construction, and analysis.

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

[4] Social desirability bias can be a concern if participants or enumerators change answers in an attempt to help.

to treatment in January 2017. 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.

Outcome variables

Following our PAP, our main analysis centers on program maize yields, total maize output, and profits. Program yields capture a direct comparison of average yield on treated farmers' enrolled land to the average yield in the control group. Since farmers rarely enroll all of their land, we might worry about various forms of spillovers across a given farmer's plots.[5] Total output should better capture the overall welfare effects for participants—net of any cross-field substitution.

To obtain "crop cut" yields, enumerators collected and physically weighed fresh and dry harvests from two randomly placed 40-square-meter boxes. Cultivated land sizes were measured by GPS readings, with enumerators walking the boundaries of each plot three times.[6] Based on the objective measures of land size and crop cut yields, we generate more standard per-acre yield and total output measures.

Enumerators collected separate crop cut measurements for enrolled land and non-enrolled land in the treatment group. While program maize yields refers to treated farmers' per-acre yield on their enrolled plot, total maize output is a weighted average of per-acre yields on the enrolled vs. non-enrolled plot (with weights proportional to the amount of land in each category). Control farmers did not separate their fields into enrolled and non-enrolled sub-plots. For them, program yields equal per-acre yields on a randomly-selected maize plot. Total output is scaled up to their total land under maize cultivation.[7]

Finally, we calculate profits as the value of output less farmers' costs. Revenues are the product of total output and average market prices from nearby vendors. For control farmers, input costs are elicited via self-reports. For treatment farmers, we know the input

[5]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.

[6]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 2019; Carletto, Savastano and Zezza 2013; Desiere and Jolliffe 2018; Gourlay, Kilic and Lobell 2017). Self-reported harvests are additionally subject to recall bias.

[7]Appendix C provides additional variable construction details.

costs on the enrolled land (since these are administered by the program); we elicit other input costs with self-reports.[8]

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 labor for land preparation, plowing, and planting). Late-season family and paid labor use, including for weeding and harvest, was collected shortly after harvest. We ascribe a monetary value to unpaid labor by calculating the mean day wage reported within the sample and devaluing this mean wage by 50% (based on rural unemployment rates in Kenyan DHS data). We then multiply this devalued mean by the number of person-days of unpaid labor.[9]

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 forced to drop observations for missing land size and harvest data. Land size data was collected during and immediately after harvest time and enumerators struggled to measure all plots of all farmers. Harvest data was recorded 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 some dropped observations, dry weight data is missing but not fresh weight. For others, we are missing both.

In Appendix E, we include a more complete description of the attrition and implement several imputation strategies to test the robustness of our results. Section 2 shows that our results are consistent across various robustness checks, including missing-data imputation methods.

Sample description and balance tests

We report summary statistics and balance tests on our pre-specified control variables in Appendix B. In addition to balance checks for treatment and control groups, we compare baseline characteristics of our two sub-populations: the "primary" sample (farmers who

[8]Overall, the data confirm that 1AF charges prices for seeds and fertilizer that are comparable to local market prices.

[9]See Appendix E for a discussion of alternative valuations of unpaid labor.

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."

Summary statistics in Table 4 confirm 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. While three-quarters of the sample used improved agricultural technologies at baseline, average input intensity is low. About half of the respondents report some knowledge of 1AF planting practices, which is primarily driven by the pre-exposed farmers.[10]

Although household size, education, and baseline maize yields differ significantly across treatment and control, an F-test of joint orthogonality of the variables in Table 4 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. 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.

We also compare participants in the primary sample to those who enrolled prior to the study (Table 7). At baseline, farmers who self-selected into early program access 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 Results

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 pre-specified controls, γ_s is a strata fixed effect to account for variations in the probability of assignment to treatment, and ε_{ics} is

[10] Appendix B provides more detail, including separate balance tables for primary and pre-exposed samples.

clustered at the level of treatment assignment (farmer group clusters).

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). [11] Otherwise, we follow the PAP closely and present results with and without controls, which does not affect our results. The standard errors in our main results are not adjusted for multiple hypothesis testing.[12]

2.1 Intermediate outcomes: behavioral change

Before turning to our main outcome variables, we examine intermediate outcomes that capture participant behavior. We would expect treatment effects to be traceable through observable changes in farmer behavior. Concretely, 1AF trains farmers on practical details such as the correct spacing between *rows of plants* and *plants within a row*, as well as the timing of different fertilizer applications.[13]

Treatment effects on production decisions

We examine several extensive- and intensive-margin aspects of production. Panel A of 1 reports treatment effects on the extensive margin of key practices, estimated with linear probability models.[14] The outcome variables in columns (1) and (2) are dummies for whether participants planted within 5 cm of 1AF's recommended spacing (between rows and for plants within rows, respectively). In column (3), fertilizer timing equals one if the participant applied fertilizer at the appropriate time of the season. 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

[11] Where appropriate, regressions also include a control for past exposure to the 1AF program. 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 (see Table 19); for parsimony we report un-weighted results.

[12] See Table 18 for MHT-adjusted and randomization inference p-values, none of which change the interpretation of our results.

[13] Each practice is tied to farming aspects that 1AF has identified as important for crop yields; these are documented in a 95-page manual. Appropriate plant density avoids plants competing for sunlight and nutrients while maximizing the use of scarce land by not planting too sparsely.

[14] See Table 11 in Appendix D for analogous results when we restrict to the primary sample.

phosphate) at planting and CAN (calcium ammonium nitrate) several weeks later. As a placebo test we include an indicator variable for participant use of an ox-driven plow. 1AF emphasizes the *timing* of plowing over the method. Since most farmers do not own an ox-plow, 1AF discourages waiting for one to become available and encourage focusing on timing relative to the rainy season onset.

Columns (1)-(3) show that treated participants are more likely than control farmers to follow spacing recommendations. Adherence to recommended spacing increases by almost 60% (160%) for row spacing (plant spacing) among farmers unacquainted with the program. Similarly, treated farmers are 170% more likely to use appropriate fertilizer timing. In column (4), we can see that the treatment did not affect the likelihood that a farmer used an ox-plow.

It is not surprising that experienced farmers may have overlooked what might seem like basic facts. 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 more on the amount of fertilizer rather than on application timing.[15]

Input intensification is an important contributor to agricultural productivity gains. Panels B and C of table 1 therefore examine program impacts on the intensive margins of production. The main variables are expenditure on fertilizer, seeds, and labor (both paid and unpaid, with the latter valued as described in section 1.3), with treatment effects on total expenditures shown in panel B and per-acre measures in panel C. Enrolled farmers spend 75% more on fertilizer, 26% more on seeds, 18% more on paid labor, and 37% more on unpaid labor. These are substantial expenditure increases, which we account for in our analysis of profit below.[16]

Comparing treatment effects for pre-exposed and new participants

Table 1 also reports estimated coefficients on previous program exposure and its interaction with the treatment. Note that these results do not reveal whether pre-exposed and new farmers differ because of past participation in the program or due to pre-existing differences.

[15]While extension manuals often distinguish between fertilizer application at planting and at top dressing, they rarely emphasize the importance of timing (e.g. National Farmers Information Services 2019).

[16]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). Cost increases are therefore unlikely driven by 1AF's input prices.

Table 1: Take-up of program practices and input use, full sample

Panel A: Program practice adoption	(1) Row Spacing	(2) Plant Spacing	(3) Fertilizer Timing	(4) Used Plow
1AF participant	0.24	0.21	0.65	0.05
	(0.040)	(0.030)	(0.030)	(0.040)
Pre-exposed	0.07	0.02	0.15	0.08
	(0.030)	(0.020)	(0.030)	(0.030)
1AF participant × pre-exposed	-0.11	-0.07	-0.18	0.03
	(0.050)	(0.040)	(0.040)	(0.040)
Control group mean	0.41	0.13	0.37	0.76
Panel B: Input costs (USD)	Fertilizer	Seeds	Paid Labor	Unpaid Labor
1AF participant	18.98	4.01	5.85	5.11
	(1.930)	(0.820)	(3.030)	(1.080)
Pre-exposed	5.62	1.23	6.17	-0.96
	(1.740)	(0.750)	(1.900)	(0.970)
1AF participant × pre-exposed	0.45	0.62	2.89	1.31
	(2.630)	(1.140)	(3.700)	(1.280)
Control group mean	25.23	15.59	31.90	13.69
Panel C: Input costs/acre	Fertilizer	Seeds	Paid Labor	Unpaid Labor
1AF participant	17.15	0.17	-4.51	1.11
	(2.750)	(1.220)	(4.300)	(2.020)
Pre-exposed	2.14	-2.00	1.44	-4.88
	(2.680)	(1.330)	(4.110)	(1.810)
1AF participant × pre-exposed	-0.73	1.41	6.53	2.23
	(3.240)	(1.470)	(5.080)	(2.360)
Control group mean	32.35	20.52	41.07	20.57
Observations	1896			

The results in Panel A of Table 1 show that the treatment effects on 1AF-promoted practices are substantially larger for new participants (85% for row spacing, 50% for plant spacing, and 38% for fertilizer timing). The pre-exposed also experience positive and significant treatment effects, ranging from 13% (14%) for row (plant) spacing to 47% for fertilizer timing. It appears that even seemingly basic information may bear repeating multiple times.

Panels B and C of Table 1 show treatment effects on per-farmer and per-acre input expenditures. Unlike the practices in panel A, the treatment effect on pre-exposed farmers is statistically indistinguishable from that on new enrollees. Combined with the fact that pre-exposed farmers in the control group use more fertilizer and paid labor than new control farmers, this may suggest that credit constraints prevent pre-exposed farmers from intensifying their production when not enrolled in the program. Pre-exposed farmers enroll slightly more land in the program than new farmers. Their higher input use could signal an intensive margin increase in maize production—or perhaps they simply had access to more land or capital. Panel C suggests that many of these differences could be driven by increases in maize acreage, although unpaid labor costs per acre are much lower for pre-exposed farmers.

2.2 Main outcomes

Next, we examine the treatment effects on our primary outcomes of interest: program maize yields, total maize output, and profits. Table 2 presents regression estimates of the average treatment effects (ATEs), following equation (1), for our three outcomes of interest. Panel A reports the results from the primary sample and panel B shows results in the full sample (Appendix tables 12 and 13 report the full set of coefficients for the regressions with control variables).

Participation in the 1AF program has an economically and statistically significant impact on maize yields, total output, and profit. The treatment effect on program yields is 25-27% across the different samples and specifications and 18-25% for total output. The estimated treatment effects on profit vary more across specifications, ranging between 9% in the full sample with covariates to 18% in the primary sample. The point estimates are attenuated in the full sample, but the effect of program participation is consistently positive and significant.

Table 2: Main results

	Program Maize Yields		Total Maize Output		Profit	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Primary sample						
1AF participant	295.43	303.99	264.90	274.15	58.02	61.93
	(37.230)	(36.810)	(85.730)	(75.820)	(30.200)	(27.520)
Baseline controls	N	Y	N	Y	N	Y
Control group mean	1128.39		1082.38		335.35	
Observations	682	682	682	682	682	682
Panel B: Full sample						
1AF participant	297.77	289.64	257.31	210.49	48.25	33.80
	(23.490)	(23.380)	(49.700)	(44.370)	(18.110)	(16.750)
Baseline controls	N	Y	N	Y	N	Y
Control group mean	1150.49		1164.00		364.61	
Observations	1896	1896	1896	1896	1896	1896

Note: This table presents results from linear regressions of the three outcomes of interest on the treatment dummy. Program maize yields are measured in kgs per acre, total maize output is measured in kgs, and profits are measured in USD. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. The primary sample includes only farmers who had never previously participated in the 1AF program. The full sample additionally includes a sample of farmers who had previously enrolled in the 1AF program. All regressions include field office (site) fixed effects. Columns 2, 4, and 6 include the full set of pre-specified controls, omitted here. 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. Regressions in Panel B in columns 2, 4, and 6 additionally control for whether households participated in 1AF in the past.

2.3 Robustness to alternative specifications and attrition

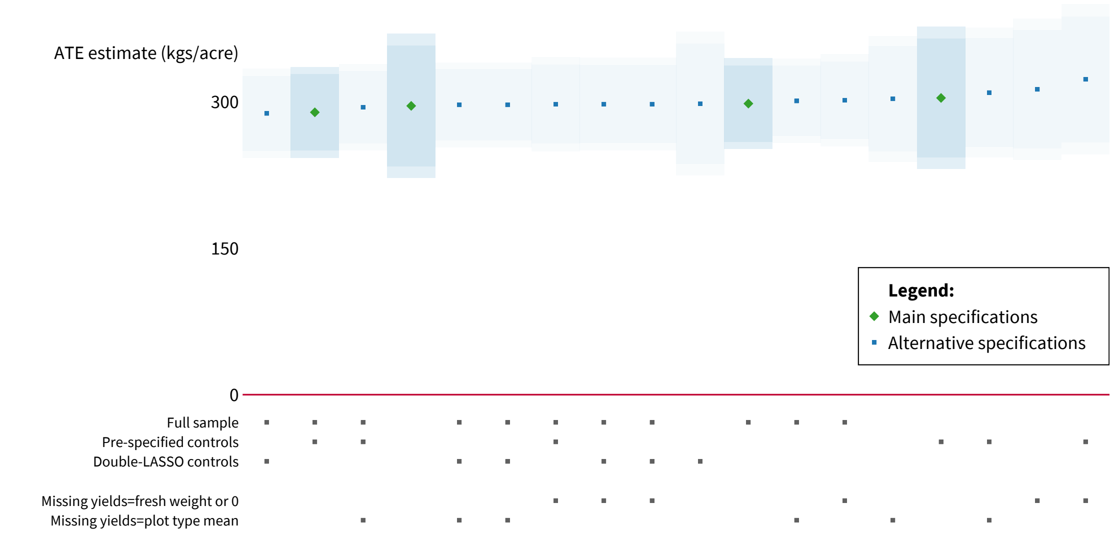
We test the robustness of our results to different specifications, approaches to dealing with attrition, and sample inclusion criteria. We describe the main variations below. We pre-specified our baseline controls, but collected many additional covariates that may improve the precision of our estimates. Following Belloni, Chernozhukov and Hansen (2014) and Urminsky, Hansen and Chernozhukov (2016), we implement a double-Lasso procedure for variable selection. In addition to pre-specified controls and fixed effects, we include additional baseline demographic, wealth, and farming experience variables.[17]

We account for attrition due to missing data in the key variables: land size and yields. This summary of our robustness checks is described in more detail in Appendix E. We impute missing values for yields and land sizes in multiple ways. For yields, we use two strategies; one imputes using non-missing "fresh weight" measurements when possible, and conservatively assumes zero yields (i.e. total crop failure) for the rest; the other replaces missing yield values with the mean yields by plot type (treated-enrolled, treated-non-enrolled, control). For land size, we use the estimated relationships in the sample of households with overlapping GPS-measured and self-reported land size to predict values for observations that are missing data. Across the different rounds of data collection, we collected several measures of self-reported land; we impute based on each measure individually and the mean of all three self-reports. For the latter, if a participant is missing one or two self-reports, we take the mean of non-missing self-reported values.

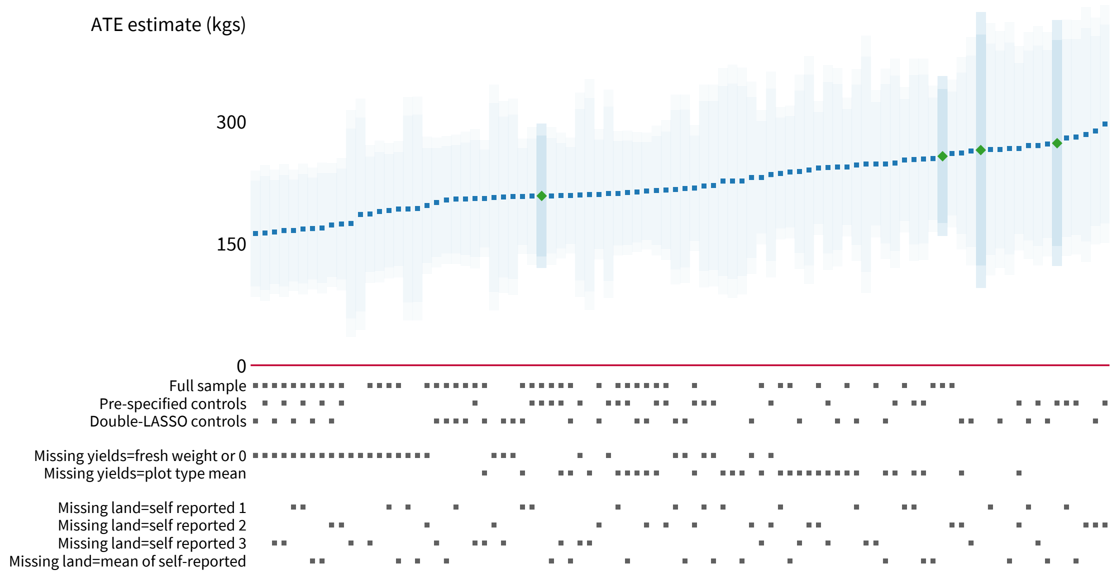
The results in Figure 1 summarize these specifications, overall revealing that our estimates are robust to different specifications, sample definitions, and attempts to deal with attrition.[18] Each vertical bar represents a different estimated ATE. Point estimates from our main specifications are indicated by green diamonds while alternative specifications are denoted by blue squares. Along the bottom, the legend indicates the corresponding specification for each estimate. For example, the left-most point estimate in Figure 1(a) is based on a regression with the full sample that includes double-LASSO controls.

[17]See e.g. Panel B of Table 7 for details about these variables.

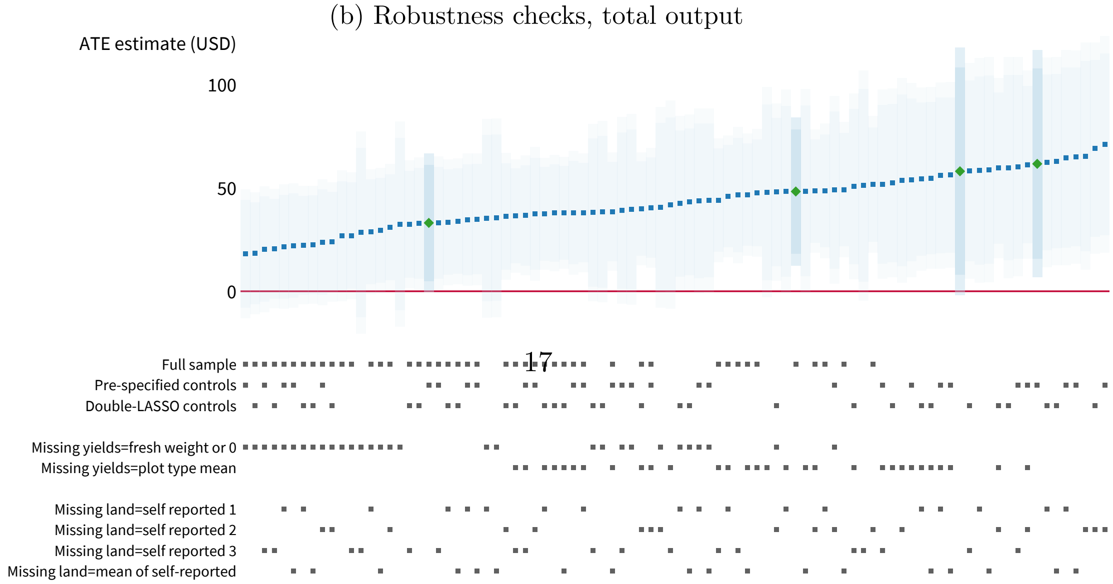
[18]Note that Figure 1 (a) has fewer permutations than (b) and (c). Since this outcome variable is based purely on crop-cut yield measurements, the various approaches to dealing with missing land measurements are irrelevant.



(a) Robustness checks, yields



(b) Robustness checks, total output



(c) Robustness checks, profit

3 Heterogeneity

We have shown that the ATE of the 1AF program is statistically and economically significant for program yields, total maize output, and profits. Does this average mask underlying heterogeneity in program impacts? Given that our profit results are slightly attenuated when we account for missing data, we may want to know whether effects are larger for specific sub-populations.

We examine treatment effect heterogeneity in three ways: unconditional quantile treatment effects, distribution regressions (Chernozhukov, Fernández-Val and Melly, 2013), and machine learning methods introduced by Chernozhukov et al. (2018). The three methods produce consistent results, so we present one set here and the others in Appendix F.

Figure 2 displays estimates of the treatment effect of program participation on unconditional quantiles of the outcome distributions (Firpo, 2007; Frölich and Melly, 2010). The treatment effect is remarkably consistent across the distribution of program maize yields. For total maize production and profits, treatment effects for farmers at the top end of the output distribution are substantially lower. For these two outcomes, we cannot detect significant treatment effects for the top two deciles of the distribution.

4 Epilogue: participation decisions a year later

We now turn to farmer behavior in the year after our study. 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.

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 re-enroll in larger proportions. If the program instead helps farmers “graduate” by nudging them across a poverty trap threshold, then we might expect fewer treated farmers to re-enroll. Column 1 of Table 3 shows that being randomly allocated to participation in 2017 did not affect the probability that farmers enrolled in 2018. We may be picking up a mix of the two aforementioned effects.

In addition to the extensive margin decision, farmers choose how much land to enroll.

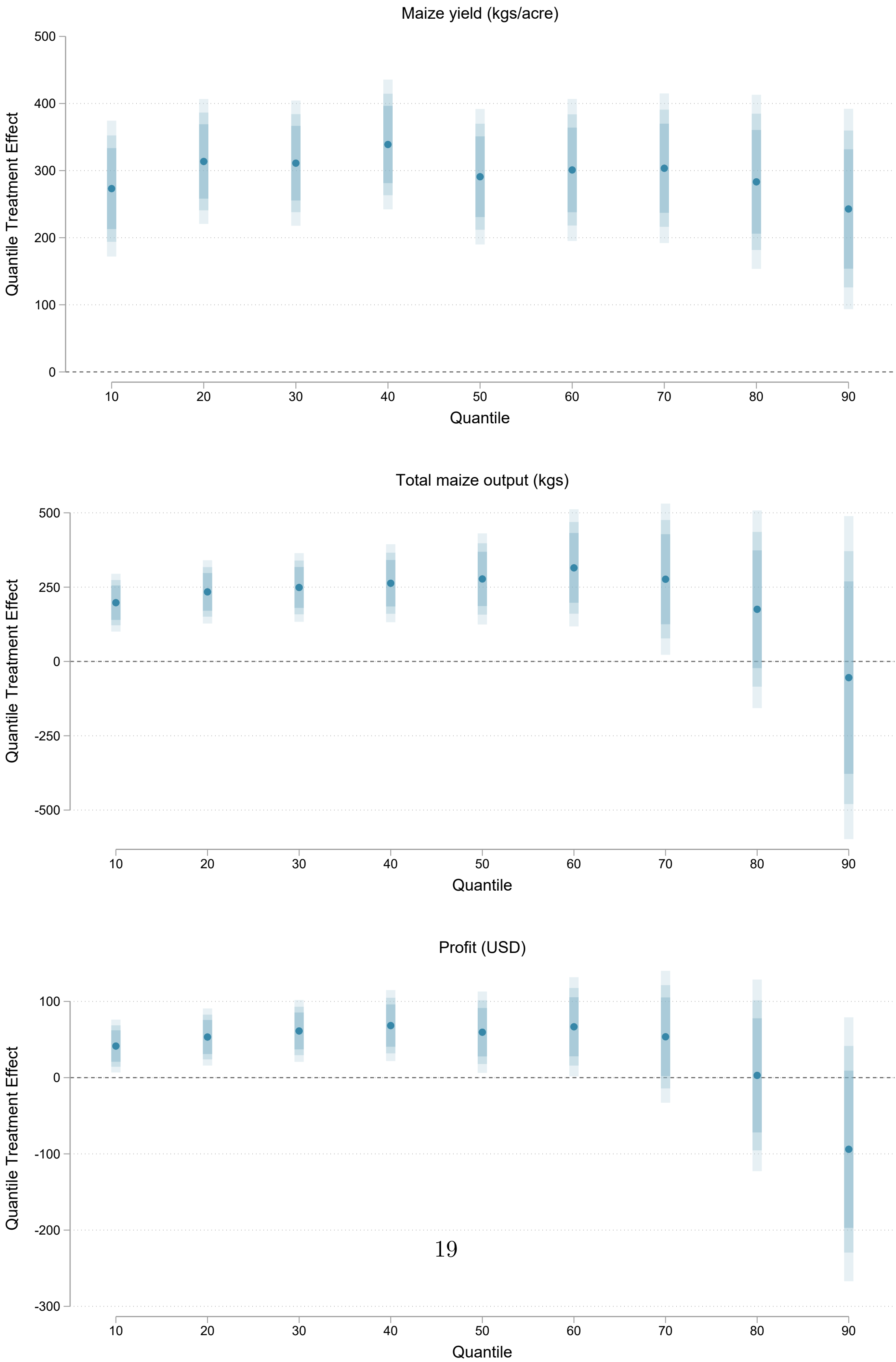


Figure 2: Unconditional quantile treatment effects

Table 3: Re-enrollment after the experiment

	Re-enrolled in 2018 (1)	Acres enrolled in 2018 (2)	Acres enrolled in 2018, missings=0 (3)
Panel A: Primary sample			
1AF participant	0.00	0.15	0.09
	(0.030)	(0.040)	(0.040)
Control group mean	0.66	0.62	0.62
Observations	682	441	682
Panel B: Full sample			
1AF participant	-0.01	0.15	0.10
	(0.020)	(0.030)	(0.030)
Control group mean	0.67	0.61	0.61
Observations	1896	1267	1896

Note: This table presents results from linear regressions of the three outcomes of interest on the treatment dummy. Re-enrolled in 2018 is a dummy variable equal to one if the client ID shows up in the 1AF administrative data as enrolled in 2018, the season after the experiment. Acres enrolled in 2018 is the land size farmers chose to enroll in the 1AF program in 2018. Column 3 assumes that any missing farmers chose to enroll zero land, and estimates the effect on acreage including those farmers. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. The primary sample includes only farmers who had never previously participated in the 1AF program. The full sample additionally includes a sample of farmers who had previously enrolled in the 1AF program. All regressions include field office (site) fixed effects. Regressions on the full sample control for whether households participated in 1AF in the past.

Larger enrollments include more credit and larger quantities of high-quality inputs, but require farmers to commit more land to a specific crop. Columns 2 and 3 of Table 3 show that farmers in the treatment group increase the amount of land that they enroll in the subsequent year. This result holds even in column 3, where we conservatively count farmers absent from the administrative data as having chosen to enroll zero acres.

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. One additional piece of evidence, found in Table 14 reveals the same pattern among pre-exposed farmers with one and two years of past program experience. Although pre-exposure is non-random, the exposure to an additional year is randomly assigned. Among these repeat-participants, we see increased enrollments of 0.23 and 0.1 acres (for farmers who had enrolled once and twice prior to the study, respectively).

5 Discussion

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 on poverty traps as well as empirical work showing that poor households benefit from bundled interventions (Banerjee et al., 2015; Bandiera et al., 2017; Balboni et al., 2020). 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.

One important question that we cannot answer with our study design is whether the bundled approach is most cost-effective, or whether a simpler program could work as well at lower cost? Our work suggests some considerations that future work on this topic could explore further. In contexts where the nature and severity of market failures vary sharply across individuals or space, it may be prohibitively costly for an organization to target interventions to each locality. 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. First, a large organization like 1AF can leverage their scale and human capital to process and distill large amounts of information 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.

A second 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. Third, 1AF 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.

Perhaps this new evidence will nudge a few cynics into reconsidering the future for (smallholder) agriculture in sub-Saharan Africa. 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, Kibrom A, Leah E M Bevis, and Christopher B Barrett. 2019. "Measurement Error Mechanisms Matter: Agricultural Intensification with Farmer Misperceptions and Misreporting." Working Paper.

Aker, Jenny C. 2011. "Dial "A" for Agriculture: A Review of Information and Communication Technologies for Agricultural Extension in Developing Countries." *Agricultural Economics*, 42(6): 631–647.

Anderson, Jock R., and Gershon Feder. 2007. "Chapter 44 Agricultural Extension." In *Handbook of Agricultural Economics*. Vol. 3 of *Agricultural Development: Farmers, Farm Production and Farm Markets*, , ed. R. Evenson and P. Pingali, 2343–2378. Elsevier.

Ashraf, Nava, James Berry, and Jesse M Shapiro. 2010. "Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia." *American Economic Review*, 100(5): 2383–2413.

Balboni, Clare, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil. 2020. "Why Do People Stay Poor?" Working Paper.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017. "Labor Markets and Poverty in Village Economies." *The Quarterly Journal of Economics*, 132(2): 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(6236): 1260799–1260799.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2013. "Profitability of Fertilizer: Experimental Evidence from Female Rice Farmers in Mali." *American Economic Review: Papers and Proceedings*, 103(3): 381–86.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2020. "Selection into Credit Markets: Evidence from Agriculture in Mali." Working Paper.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on Treatment Effects after Selection among High-Dimensional Controls." *The Review of Economic Studies*, 81(2): 608–650.

Birkhaeuser, Dean, Robert E. Evenson, and Gershon Feder. 1991. "The Economic Impact of Agricultural Extension: A Review." *Economic Development and Cultural Change*, 39(3): 607–650.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. 2005. "Distributional Impacts of the Self-Sufficiency Project." National Bureau of Economic Research Working Paper 11626.

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

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

Bravo-Ortega, Claudio, and Daniel Lederman. 2005. "Agriculture and National Welfare around the World: Causality and International Heterogeneity since 1960." The World Bank 3499.

Byerlee, Derek, Alain de Janvry, and Elisabeth Sadoulet. 2009. "Agriculture for Development: Toward a New Paradigm." *Annual Review of Resource Economics*, 1(1): 15–31.

Carletto, Calogero, Sara Savastano, and Alberto Zezza. 2013. "Fact or Artifact: The Impact of Measurement Errors on the Farm Size–Productivity Relationship." *Journal of Development Economics*, 103: 254–261.

Carter, Michael R., Emilia Tjernström, and Patricia Toledo. 2019. "Heterogeneous Impact Dynamics of a Rural Business Development Program in Nicaragua." *Journal of Development Economics*, 138: 77–98.

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly. 2013. "Inference on Counterfactual Distributions." *Econometrica*, 81(6): 2205–2268.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val. 2018. "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments." National Bureau of Economic Research Working Paper 24678.

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

Collier, Paul, and Stefan Dercon. 2014. "African Agriculture in 50 Years: Smallholders in a Rapidly Changing World?" *World Development*, 63(C): 92–101.

Desiere, Sam, and Dean Jolliffe. 2018. "Land Productivity and Plot Size: Is Measurement Error Driving the Inverse Relationship?" *Journal of Development Economics*, 130: 84–98.

Dubey, Priyanka, and Rosaine 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." International Initiative for Impact Evaluation Field Report.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2008. "How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya." *American Economic Review: Papers and Proceedings*, 98(2): 482–488.

Feder, Gershon, Richard E. Just, and David Zilberman. 1985. "Adoption of Agricultural Innovations in Developing Countries: A Survey." *Economic Development and Cultural Change*, 33(2): 255–298.

Firpo, Sergio. 2007. "Efficient Semiparametric Estimation of Quantile Treatment Effects." *Econometrica*, 75(1): 259–276.

Frölich, Markus, and Blaise Melly. 2010. "Estimation of Quantile Treatment Effects with Stata." *The Stata Journal*, 10(3): 423–457.

Gollin, Douglas, Casper Worm Hansen, and Asger Wingender. 2018. "Two Blades of Grass: The Impact of the Green Revolution." National Bureau of Economic Research Working Paper 24744.

Gourlay, Sydney, Talip Kilic, and David Lobell. 2017. "Could the Debate Be Over? Errors in Farmer-Reported Production and Their Implications for the Inverse Scale-Productivity Relationship in Uganda." The World Bank 8192.

Government of Kenya. 2010. "Agricultural Sector Development Strategy 2010-2020." Technical Report.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. 2014. "Learning Through Noticing: Theory and Evidence from a Field Experiment." *The Quarterly Journal of Economics*, 129(3): 1311–1353.

Heckman, James J., Jeffrey Smith, and Nancy Clements. 1997. "Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts." *The Review of Economic Studies*, 64(4): 487–535.

Kirimi, Lilian, Nicholas Sitko, Thomas S. Jayne, Francis Karin, Milu Muyanga, Megan Sheahan, James Flock, and Gilbert Bor. 2011. "A Farm Gate-to-Consumer Value Chain Analysis of Kenya's Maize Marketing System." Tegemeo Institute of Agricultural Policy and Development Technical Report WPS 44/2011.

Ligon, Ethan A., and Elisabeth Sadoulet. 2011. "Estimating the Effects of Aggregate Agricultural Growth on the Distribution of Expenditures." Department of Agricultural & Resource Economics, UC Berkeley 1115.

Magruder, Jeremy R. 2018. "An Assessment of Experimental Evidence on Agricultural Technology Adoption in Developing Countries." *Annual Review of Resource Economics*, 10(1): 299–316.

Matsuyama, Kiminori. 1992. "Agricultural Productivity, Comparative Advantage, and Economic Growth." *Journal of Economic Theory*, 58(2): 317–334.

McArthur, John W., and Gordon C. McCord. 2017. "Fertilizing Growth: Agricultural Inputs and Their Effects in Economic Development." *Journal of Development Economics*, 127: 133–152.

National Farmers Information Services. 2019. "Field Management – NAFIS." *http://www.nafis.go.ke/agriculture/maize/field-management-practices/.*

One Acre Fund. 2020. “How We Grow.” *https://oneacrefund.org/what-we-do/how-we-grow/.*

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

Sheahan, Megan, Roy 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, Tavneet. 2011. “Selection and Comparative Advantage in Technology Adoption.” *Econometrica*, 79(1): 159–209.

Tarozzi, Alessandro, Aprajit Mahajan, Brian Blackburn, Dan Kopf, Lakshmi Krishnan, and Joanne Yoong. 2014. “Micro-Loans, Insecticide-Treated Bednets, and Malaria: Evidence from a Randomized Controlled Trial in Orissa, India.” *American Economic Review*, 104(7): 1909–1941.

Tinsley, Elaine, and Natalia Agapitova. 2018. “Private Sector Solutions to Helping Smallholders Succeed: Social Entreprise Business Models in the Agriculture Sector.” The World Bank Report.

Tjernström, Emilia, Michael R. Carter, and Travis Lybbert. 2018. “The Dirt on Dirt: Soil Characteristics and Variable Fertilizer Returns in Kenyan Maize Systems.” Working Paper.

Udry, Christopher, Federica di Battista, Mathias Fosu, Markus Goldstein, Alev Gurbuz, Dean Karlan, and Shashidhara Kolavalli. 2019. “Information, Market Access and Risk: Addressing Constraints to Agricultural Transformation in Northern Ghana.” Draft Report.

Urminsky, Oleg, Christian Hansen, and Victor Chernozhukov. 2016. “Using Double-Lasso Regression for Principled Variable Selection.” Working Paper.

World Bank. 2008. “World Development Report 2008: Agriculture for Development.” World Bank Report.

A A – Pre-Analysis Plan

International Initiative for Impact Evaluation

New Delhi, 23rd July 2018

To whom it may concern

Reg.: Confirmation of the review of the Pre-Analysis Plan (PAP)

Dear Sir/Madam

This letter is to confirm that 3ie reviewed the PAP of the Impact Evaluation of the One Acre Fund program on yields and profits of maize and beans farmers in Teso, Kenya.

The PAP was submitted to 3ie by Maya Duru and Kim Siegal. The PAP review process was led by Rosaine N. Yegbemey between July and November, 2016.

The PAP went through three main rounds of review with several iterations of comments and a couple of Skype calls. Considering the context of the evaluation and the 1AF team's responses to the comments, the revised PAP was found to be appropriate to the goals of the study and of sufficient level or rigor.

For any questions on the PAP review process, please contact Rosaine N. Yegbemey at ryegbemey@3ieimpact.org.

Marie Gaarder

Marie Gaarder
Director of Evaluation Office and Global Director for Innovation and Country Engagement
International Initiative for Impact Evaluation (3ie)

New Delhi
202–203, Rectangle One
D-4, Saket District Centre
New Delhi – 110017, India
3ie@3ieimpact.org
Tel: +91 11 4989 4444

London
c/o LIDC, 36 Gordon Square,
London WC1H 0PD
United Kingdom
3ieuk@3ieimpact.org
Tel: +44 207 958 8351/8350

Washington, DC
1029 Vermont Avenue, NW, Suite 1000
Washington, DC 20005
United States of America
3ieus@3ieimpact.org
Tel: +1 202 629 3939

B B – Additional summary statistics and balance checks

B.1 Balance by treatment assignment

Table 4: 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.

Table 5: 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 6: 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.

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

Table 7: 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.

B.2.1 Distributions of nonbinary baseline variables by treatment status

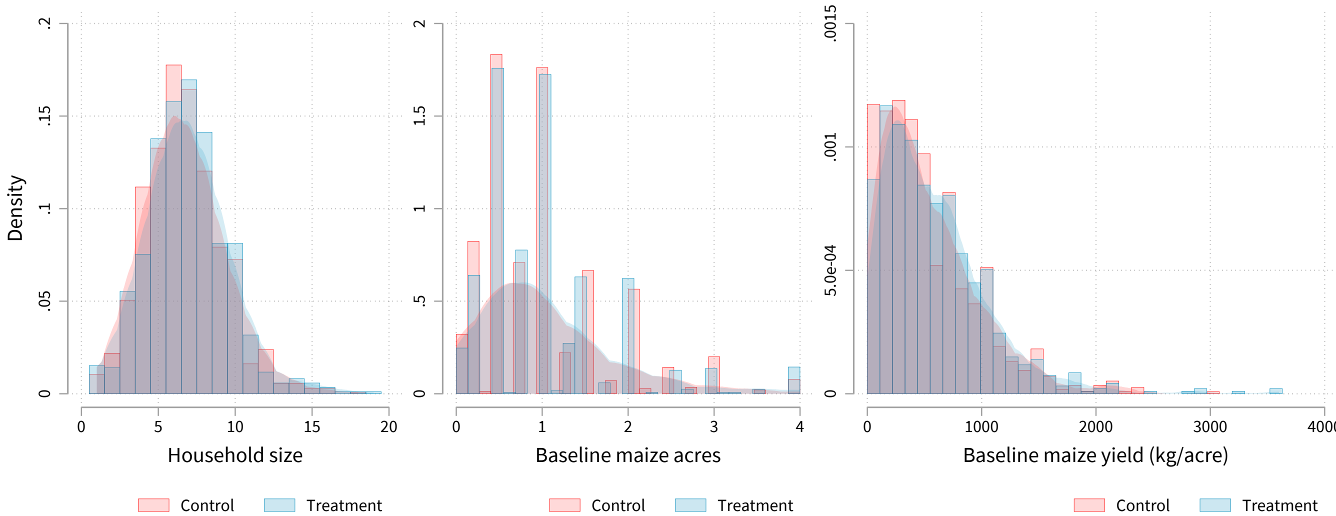


Figure 3: Distributions of non-binary baseline characteristics, by treatment status

C C – 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. For treated farmers, when we consider farmer-level outcomes, we average yields across enrolled and non-enrolled land, weighted by the proportion of land farmers enrolled.

We calculate projected revenues using average market prices from nearby vendors covering post-harvest months, multiplied by 1.08 to account for typical price increases over the consumption/selling season. We calculate farmer 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. Profit is simply the difference between projected farmer revenues and costs.

Note that in the original PAP, all labor costs were specified to be devalued by 50%. We feel that it is more appropriate to only make this correction for unpaid labor, as this more appropriately reflects the expected wage a household laborer could earn in the market. We show in Table 8 that our profit results are robust to valuing all labor at the market rate, although unsurprisingly the treatment effect does decrease slightly in magnitude.

Panel B of Table 11 (and in 1 in the main text) considers input use valued in USD, while tables 9 and 10 measure fertilizer use in kilograms. We detect a sizeable increase in fertilizer use when measured in kilograms, reinforcing 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 Panel A of Table 11. Farmers in the treatment group are not only using more fertilizer at the "correct" time, but also using less fertilizer at incorrect times.

Table 8: Profit with different labor cost definitions,

	Primary Sample		Full Sample	
	(1)	(2)	(3)	(4)
	Profit (PAP def)	Profit (Mkt rate labor)	Profit (PAP def)	Profit (Mkt rate labor)
1AF participant	58.021*	53.122*	48.245***	42.297**
	(30.196)	(30.019)	(18.105)	(17.933)
Observations	682	682	1896	1896
R^2	0.124	0.120	0.075	0.074
Control Mean Dep. Var	335.349	320.703	364.615	350.920

This table presents results from linear regressions of the outcomes in each column on the treatment dummy. Columns (1) and (3) value paid labor at market rates and values unpaid labor at 50% of market rate (following PAP). Columns (2) and (4) value all labor at market rates. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects. Full sample regressions additionally control for pre-exposure status.

Table 9: 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	-7.169***	15.472***
	(1.906)	(0.145)	(1.846)	(2.378)
Observations	682	682	682	682
R^2	0.236	0.013	0.221	0.211
Control Mean Dep. Var	8.976	0.213	10.256	18.068

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 10: 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	-7.188***	15.821***
	(1.995)	(0.177)	(1.879)	(2.325)
Past 1AF participant=1	7.708***	0.235	-2.763*	5.281**
	(1.650)	(0.261)	(1.531)	(2.065)
1AF participant=1 × Past 1AF participant=1	-2.773	0.258	3.497*	0.351
	(2.555)	(0.326)	(1.813)	(3.020)
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	21.945

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.

D D – Intermediate outcomes under alternative sample specifications

Table 11: Take-up of program practices and input use, primary sample

Panel A:	(1)	(2)	(3)	(4)
Take-up of program practices	Row Spacing	Plant Spacing	Fertilizer Timing	Used Plow
1AF participant	0.22***	0.20***	0.66***	0.05
	(0.040)	(0.030)	(0.030)	(0.040)
Control group mean	0.37	0.09	0.26	0.73
N	682	682	682	682
Panel B:				
Input costs (USD)	Fertilizer	Seeds	Paid Labor	Unpaid Labor
1AF participant	18.74***	3.84***	5.32*	4.90***
	(1.950)	(0.810)	(2.920)	(1.000)
Control group mean	20.55	14.77	29.63	14.65
N	682	682	682	682
Panel C:				
-2 *Input costs/acre*	Fertilizer	Seeds	Paid Labor	Unpaid Labor
1AF participant	17.02***	0.21	-4.09	0.61
	(2.720)	(1.110)	(3.770)	(1.960)
Control group mean	28.34	21.18	41.85	23.48
N	682	682	682	682

This table presents results from linear regressions of the outcome variables in each column on the treatment dummy. Panel A shows the effect of 1AF participation on the use of practices recommended by the NGO: the spacing used between plants and fertilizer use at the correct time in the season. 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). Used plow is an indicator for whether or not the farmer used a plow to prepare their plot. Panel B shows the effect of 1AF participation on the intensive margin of farmer expenses for fertilizer, seeds, paid labor, and unpaid labor. Costs are expressed in USD. For more on how we define labor costs, see Appendix C. Panel C shows the effect of 1AF participation on investment per acre for the cost types shown in Panel B. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects. This table includes only farmers in the primary sample.

Note: This table presents results from linear regressions of the outcome variables in each column on the treatment dummy. Panel A shows the effect of 1AF participation on the use of practices recommended by the NGO: the spacing used between plants and fertilizer use at the correct time in the season. 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). Used plow is an indicator for whether or not the farmer used a plow to prepare their plot. Panel B shows the effect of 1AF participation on the intensive margin of farmer expenses for fertilizer, seeds, paid labor, and unpaid labor. Costs are expressed in USD. For more on how we define labor costs, see Appendix C. Panel C shows the effect of 1AF participation on investment per acre for the cost types shown in Panel B. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects. This table includes only farmers in the primary sample.

Table 12: Primary outcomes with controls, primary sample

	(1) Program Maize Yields	(2) Total Maize Output	(3) Profit
1AF participant	303.99***	274.15***	61.93**
	(36.81)	(75.82)	(27.52)
Married	119.74*	-8.55	-5.14
	(65.30)	(147.30)	(53.60)
Household head has secondary school	4.52	262.50***	88.15***
	(50.68)	(88.32)	(32.47)
Household income >50% from farm labor	28.20	47.94	21.24
	(46.10)	(79.68)	(29.37)
Used improved ag technology in 2016	-99.26**	-146.89*	-62.64*
	(47.00)	(86.66)	(31.70)
Reports knowledge of 1AF practices	22.78	52.70	2.10
	(74.40)	(144.61)	(53.95)
Intercropped maize and beans in 2016	-21.71	-70.62	-23.67
	(47.53)	(92.07)	(33.85)
Reports having credit access in 2016	-93.83**	-65.93	-26.95
	(37.53)	(79.57)	(28.89)
Household size	6.09	29.10*	11.42**
	(7.66)	(15.60)	(5.45)
Acres under maize cultivation in 2016	72.03**	362.62***	111.05***
	(29.22)	(73.38)	(26.66)
Maize yield (kg/acre) in 2016	0.18***	0.45***	0.15***
	(0.05)	(0.10)	(0.04)
FAW incidence	-70.60	-135.38	-50.90
	(45.96)	(117.44)	(43.37)
Observations	682	682	682
R^2	0.199	0.254	0.211

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

Table 13: Primary outcomes with controls, full sample

	(1) Program Maize Yields	(2) Total Maize Output	(3) Profit
1AF participant	289.64***	210.49***	33.80**
	(23.38)	(44.37)	(16.75)
Past 1AF participant	-23.53	26.84	10.46
	(28.62)	(65.17)	(23.35)
Married	60.66*	59.95	11.68
	(36.26)	(69.44)	(25.88)
Household head has secondary school	9.78	187.58***	62.72***
	(26.20)	(51.14)	(18.66)
Household income >50% from farm labor	56.92**	135.18**	54.96***
	(26.22)	(51.93)	(19.36)
Used improved ag technology in 2016	-42.32	-72.97	-34.45
	(36.54)	(65.33)	(24.13)
Reports knowledge of 1AF practices	31.68	79.06	20.76
	(30.92)	(55.74)	(21.09)
Intercropped maize and beans in 2016	-13.67	-16.00	1.83
	(26.04)	(56.03)	(20.76)
Reports having credit access in 2016	-21.37	-27.78	-19.44
	(24.58)	(47.10)	(17.47)
Household size	4.47	38.72***	13.30***
	(4.60)	(9.76)	(3.56)
Acres under maize cultivation in 2016	30.60*	374.50***	118.11***
	(16.78)	(54.46)	(19.89)
Maize yield (kg/acre) in 2016	0.13***	0.36***	0.11***
	(0.03)	(0.08)	(0.03)
FAW incidence	-28.56	-73.86	-27.01
	(31.57)	(69.45)	(25.24)
Observations	1896	1896	1896
R^2	0.165	0.215	0.168

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

Table 14: Acres enrolled in 2018, from past new enrollees

	(1) 2018 Acres (Joined 2016)	(2) 2018 Acres (Joined 2015)
1AF participant	0.23***	0.10**
	(0.04)	(0.04)
Observations	344	262
R^2	0.177	0.143
Control Mean Dep. Var	0.58	0.63

Results in this table are from linear regressions of the outcome variables on the treatment dummy. The outcome variable 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.

E E – Robustness and external validity

E.1 Distributions of outcome variables

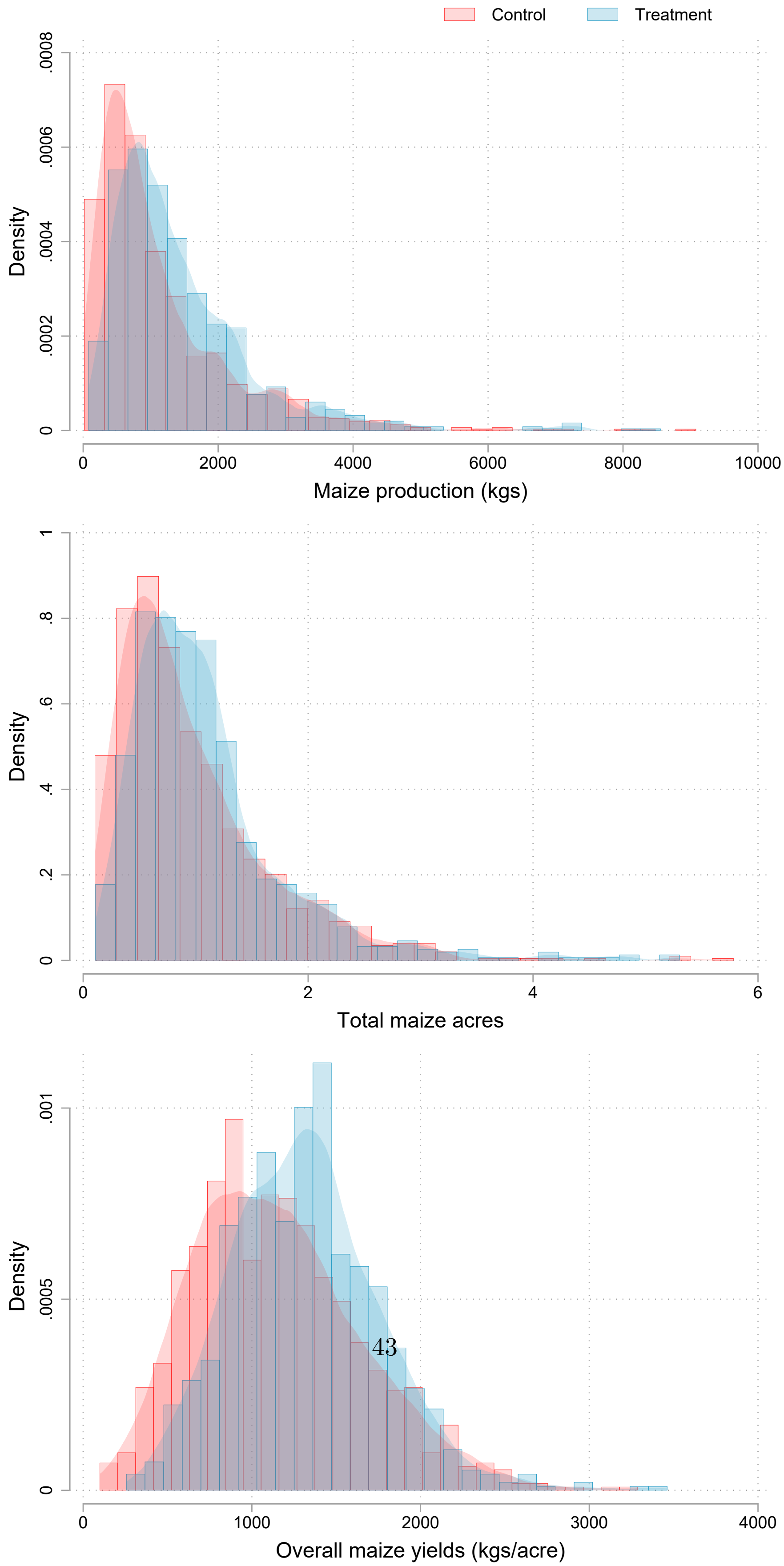
Some readers may wish to see what the raw data look like, in addition to the regression results. Figure 4 shows the distributions of total production, maize acreage, and overall maize yields by treatment status. For treatment farmers, total maize production and acreage include both enrolled and non-enrolled plots, and the yields represent total harvests divided by each plot's respective size. In the top figure, we can see that treatment farmers are less likely to attain very low production levels. Further, the treatment group distribution lies to the right of the control group distribution for much of the support.

The second panel shows that treatment farmers also plant slightly more land to maize (by 0.1 acres on average, significant at the 1% level). The bottom panel shows overall per-farmer yields. Similar to the level effects, treatment group farmers are less likely to experience very low yields, and the distribution of yields is clearly shifted to the right. The top figure here corresponds to what we call the total output measure, since it accounts both for yields on the enrolled plot and the non-enrolled plot. The bottom panel is different than the program maize yields as it counts yields on both program and non-program land for treated farmers.

E.2 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 E.2.1.

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


Control
Treatment
Density
Maize production (kgs)
Density
Total maize acres
Density
Overall maize yields (kgs/acre)

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.

E.2.1 Attrition prior to season and external validity

Table 15 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 Tegemeo sample (see Figure 5).[19] Note that the Tegemeo survey relies on self-reported acreage; we therefore report our baseline self-reported acreage variable (dotted line in Figure 5) in addition to the GPS-measured land sized (short-dashed line in Figure 5). 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 16 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 practices, 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.

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

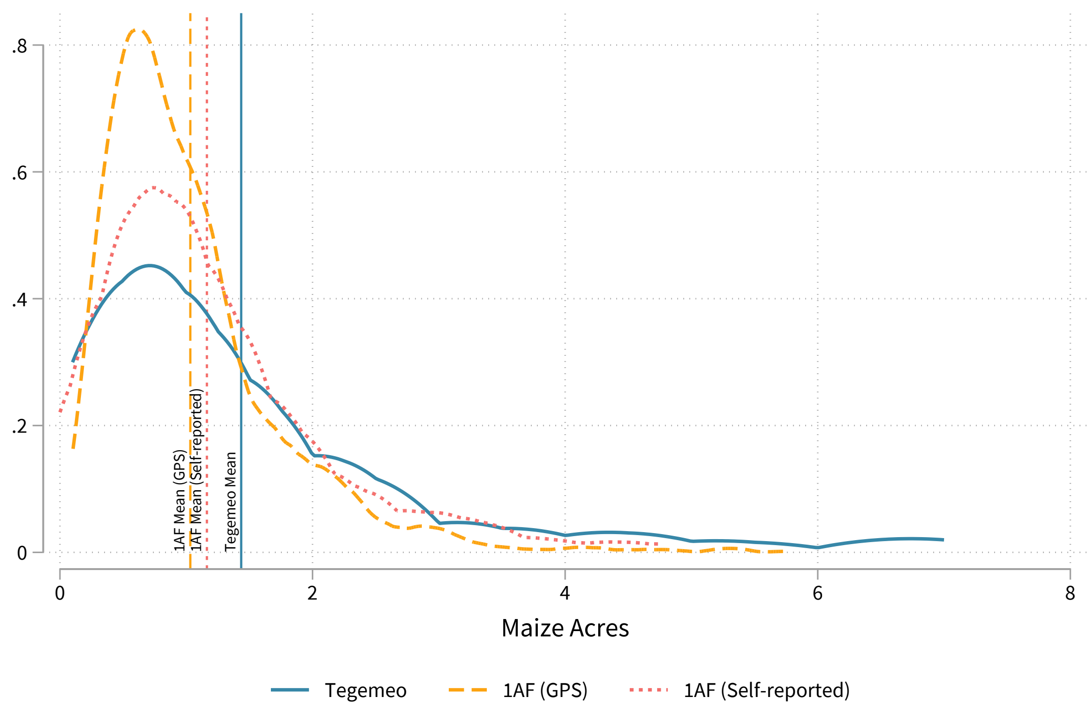


Figure 5: Comparison of maize acres across study sample and TAPRA survey

Table 15: 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

Table 16: 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

E.2.2 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.

For yields, we use two different imputation strategies. 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 use 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.

E.3 Alternative specification details

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.[20] Following Belloni, Chernozhukov and Hansen (2014) and Urminsky, Hansen and Chernozhukov (2016), we implement a double-Lasso 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 17, demonstrate that the results are not particularly sensitive to the inclusion of an optimal set of controls.

[20] 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 17: Main outcomes with double-Lasso variable selection

	Program Maize Yields			Total Maize Output			Profit		
	(1)	(2)	(3)	(4)	(5)	(6)	(8)	(8)	(9)
Panel A: Primary sample									
1AF participant	295.43***	303.99***	299.67***	264.90***	274.15***	266.08***	58.02*	61.93**	59.69**
	(37.230)	(36.810)	(37.190)	(85.730)	(75.820)	(73.720)	(30.200)	(27.520)	(27.070)
Pre-specified controls	N	Y	N	N	Y	N	N	Y	N
Lasso-selected controls	N	N	Y	N	N	Y	N	N	Y
Control group mean		1128.39			1082.38			335.35	
Observations		682			682			682	
Panel B: Full sample									
1AF participant	297.77***	289.64***	290.76***	257.31***	210.49***	206.32***	48.25***	33.80**	32.90**
	(23.490)	(23.380)	(23.190)	(49.700)	(44.370)	(41.300)	(18.110)	(16.750)	(15.860)
Pre-specified controls	N	Y	N	N	Y	N	N	Y	N
Lasso-selected controls	N	N	Y	N	N	Y	N	N	Y
Control group mean		1150.49			1164.00			364.61	
Observations		1896			1896			1896	

Note: This table presents results from linear regressions of the three outcomes of interest on the treatment dummy. Program maize yields are measured in kgs per acre, total maize output is measured in kgs, and profits are measured in USD. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. The primary sample includes only farmers who had never previously participated in the 1AF program. The full sample additionally includes a sample of farmers who had previously enrolled in the 1AF program. All regressions include field office (site) fixed effects. Columns 2, 5, and 8 include the full set of pre-specified controls, omitted here. 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. Regressions on the full sample additionally control for whether households participated in 1AF in the past. Columns 3, 6, and 9 use the double-Lasso approach of Belloni, Chernozhukov and Hansen (2014); Urminsky, Hansen and Chernozhukov (2016) to select control variables using 10-fold cross validation. The Lasso regressions select from all baseline covariates included in columns 2, 5, and 8, 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 18: Main outcomes with MHT-adjusted p-values

	Program Maize Yields		Total Maize Output		Profit	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Primary sample						
1AF participant	295.43	303.99	264.90	274.15	58.02	61.93
Unadjusted p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.06)	(0.03)
Westfall-Young adjusted p-value	(0.00)	(0.00)	(0.01)	(0.00)	(0.09)	(0.05)
Holm-Bonferroni adjusted p-value	(0.00)	(0.00)	(0.01)	(0.00)	(0.06)	(0.03)
Randomization Inference p-value	(0.00)	(0.00)	(0.01)	(0.00)	(0.06)	(0.03)
Baseline controls	N	Y	N	Y	N	Y
Control group mean	1128.39		1082.38		335.35	
Observations	682	682	682	682	682	682
Panel B: Full sample						
1AF participant	297.77	289.64	257.31	210.49	48.25	33.80
Unadjusted p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.05)
Westfall-Young adjusted p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.07)
Holm-Bonferroni adjusted p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.05)
Randomization Inference p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.05)
Baseline controls	N	Y	N	Y	N	Y
Control group mean	1150.49		1164.00		364.61	
Observations	1896	1896	1896	1896	1896	1896

Note: This table presents results from linear regressions of the three outcomes of interest on the treatment dummy. Westfall-Young and Holm-Bonferroni adjusted p-values are estimated with 10000 bootstrap replications, with adjustments done considering each sample definition and regressions with/without controls as separate families of outcomes. Randomization Inference p-values are estimated using 2000 re-randomizations. Program maize yields are measured in kgs per acre, total maize output is measured in kgs, and profits are measured in USD. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. The primary sample includes only farmers who had never previously participated in the 1AF program. The full sample additionally includes a sample of farmers who had previously enrolled in the 1AF program. All regressions include field office (site) fixed effects. Columns 2, 4, and 6 include the full set of pre-specified controls, omitted here. 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. Regressions in Panel B in columns 2, 4, and 6 additionally control for whether households participated in 1AF in the past.

Table 19: Main outcomes with MHT-adjusted p-values and spillover weights

	Program Maize Yields		Total Maize Output		Profit	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Primary sample						
1AF participant	293.20	299.46	265.21	268.19	60.00	60.79
Unadjusted p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.06)	(0.03)
Westfall-Young adjusted p-value	(0.00)	(0.00)	(0.01)	(0.01)	(0.09)	(0.07)
Holm-Bonferroni adjusted p-value	(0.00)	(0.00)	(0.01)	(0.00)	(0.06)	(0.03)
Randomization Inference p-value	(0.00)	(0.00)	(0.01)	(0.00)	(0.06)	(0.03)
Baseline controls	N	Y	N	Y	N	Y
Control group mean	1128.39		1082.38		335.35	
Observations	682	682	682	682	682	682
Panel B: Full sample						
1AF participant	294.13	291.70	233.27	212.24	41.33	34.53
Unadjusted p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.04)
Westfall-Young adjusted p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)	(0.06)
Holm-Bonferroni adjusted p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.04)
Randomization Inference p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.04)
Baseline controls	N	Y	N	Y	N	Y
Control group mean	1150.49		1164.00		364.61	
Observations	1896	1896	1896	1896	1896	1896

Note: This table presents results from linear regressions of the three outcomes of interest on the treatment dummy. Regressions are weighted using the inverse of the probability of exposure to treated farmers. Westfall-Young and Holm-Bonferroni adjusted p-values are estimated with 10000 bootstrap replications, with adjustments done considering each sample definition and regressions with/without controls as separate families of outcomes. Randomization Inference p-values are estimated using 2000 re-randomizations. Program maize yields are measured in kgs per acre, total maize output is measured in kgs, and profits are measured in USD. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. The primary sample includes only farmers who had never previously participated in the 1AF program. The full sample additionally includes a sample of farmers who had previously enrolled in the 1AF program. All regressions include field office (site) fixed effects. Columns 2, 4, and 6 include the full set of pre-specified controls, omitted here. 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. Regressions in Panel B in columns 2, 4, and 6 additionally control for whether households participated in 1AF in the past.

F F – Additional heterogeneity results

Beyond the quantile regression results found in Section 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

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

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 6 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.

We can see that the effect of the treatment on the proportion of participants above a certain yield is substantial, across the bulk of the control group yield distribution. The results for maize output and profits both suggest that the effects are largest around the lower end of the distribution and they attenuate at large values of the outcomes. While total maize output can at least in theory be expanded by bringing more land under cultivation, maize yields likely have some physiological upper bound, beyond which decreasing marginal returns to inputs start to make additional intensification unprofitable.

Figure 7 implements methods proposed by Chernozhukov et al. (2018) to estimate key features of heterogeneous effects on our outcomes of interest. A key difference between this approach and the previous two is that it focuses on whether specific covariates can predict the size of participants' treatment effects. Figure **??** 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 be that we are not including the correct covariates.

Our discussion of below draws heavily on the discussion in Chernozhukov et al. (2018). Section 6.2 in their paper describes the implementation algorithm in detail. A key challenge with machine learning tools in high-dimensional settings is that they typically require strong assumptions to produce consistent estimators of conditional average treatment effects (CATE). The new method developed in Chernozhukov et al. (2018) sacrifices some generalizability, but in return the authors are able to rely on fewer assumptions.

In particular, instead of trying to make inference on the full CATE function, the method focuses on making inference on key *features* of the CATE. These features are (1) the Best Linear Predictor (BLP) of the CATE function, (2) Sorted Group Average Treatment Effects (GATES), reporting predicted treatment effects at different deciles of the predicted treatment effect distribution, and (3) Classification Analysis (CLAN), showing how covariates of interest differ between the units that we predict will be the most and least affected, and these most-affected and least-affected groups are also defined by the highest and lowest deciles of the predicted treatment effect distribution. Below, we provide some more intuition for the BLP, as it is perhaps the least obvious of the three (BLP results available from the authors).

Briefly, the method splits the data into an auxiliary subset, separate from the main data (the data is split into main and auxiliary many times, as is standard with ML techniques). Letting Y^0 and Y^1 denote potential outcomes under control and treatment, respectively, we can write out two key functions: $b_0(Z) := E[Y^0|Z]$, which is the baseline conditional average, and $s_0(Z) := E[Y^1|Z] - E[Y^0|Z]$. Given a randomly assigned treatment variable D, a known propensity score $p(Z)$, and a few more assumptions on the propensity score, the observed outcome can be written as a regression function (here, conditional on D, Z): $Y = b_0(Z) + Ds_0(Z) + U$, where $E[U|Z, D] = 0$.

We then proceed by using each auxiliary sample to train an ML estimator and obtain ML estimates of the baseline and treatment effects, called proxy scores. We will refer to the estimated proxies of $b_0(Z)$ and $s_0(Z)$ as $B(Z)$ and $S(Z)$, respectively. Note that we can then use these predicted proxies in the main sample to estimate the BLP of the conditional average treatment effect. Essentially, we regress the observed outcome on the treatment variable minus the propensity score (to estimate the average treatment effect, or ATE), and on the treatment variable minus the propensity score *interacted with* deviations of

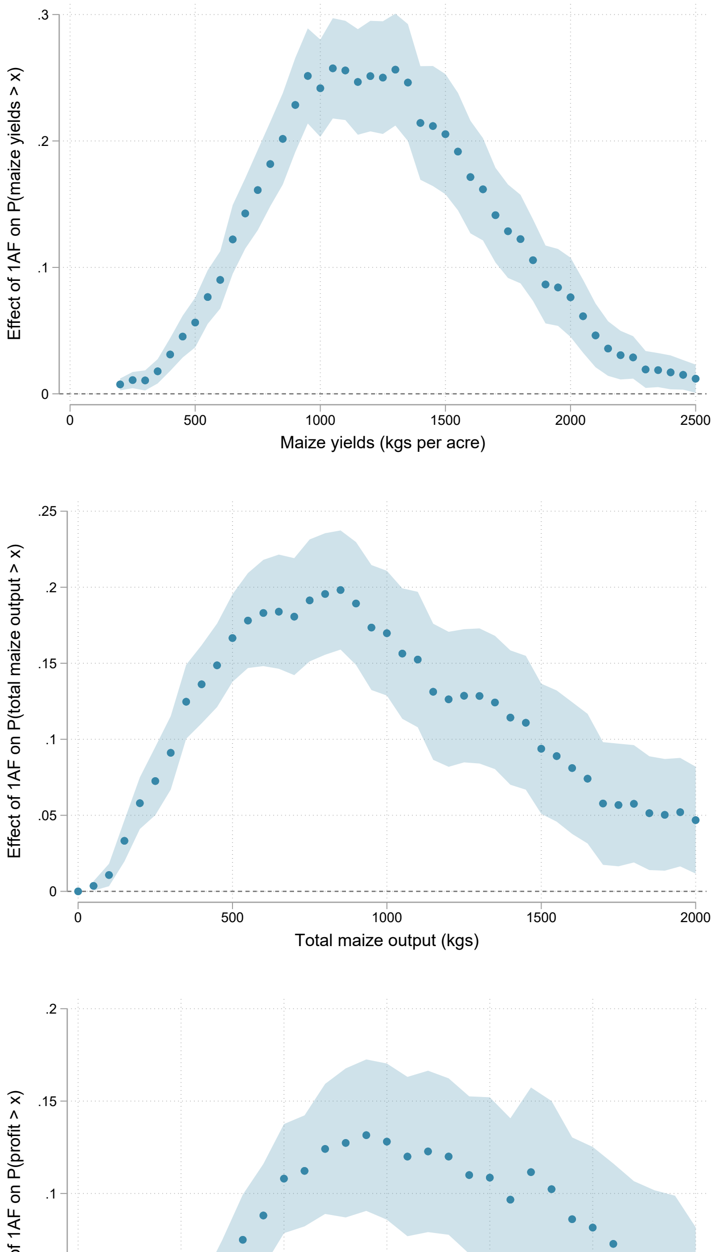
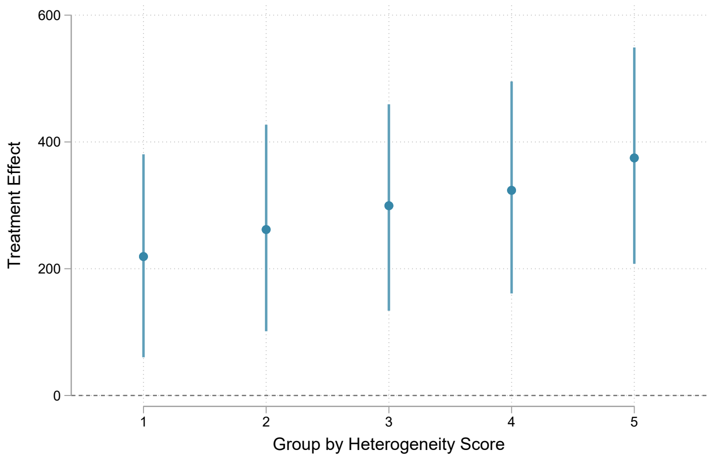
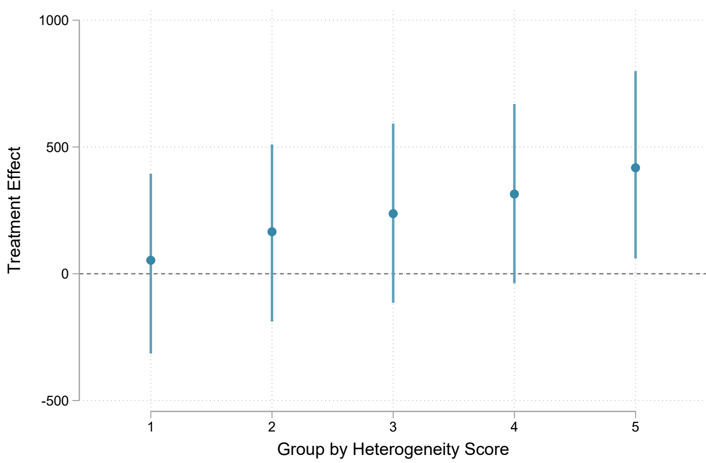


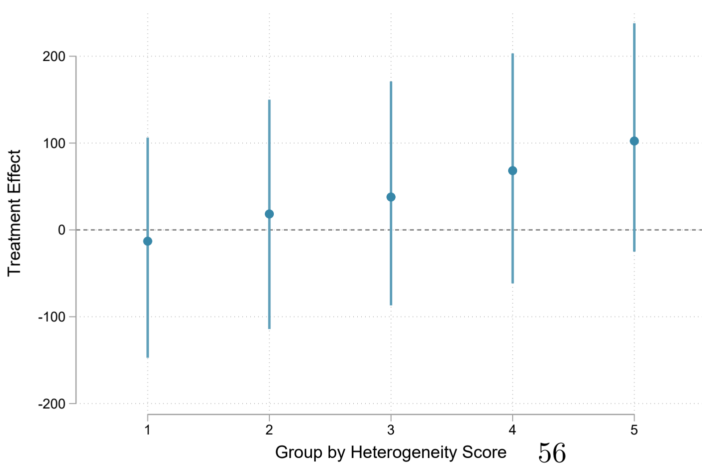
Figure 6: Distribution regression results



(a) Yields



(b) Total output



(c) Profits

Figure 7: Sorted Group Average Treatment Effects, estimated with Neural Nets

the $S(Z)$ that we estimated in the auxiliary data from the expected value of $S(Z)$ in the main sample. The coefficient on this second interaction term is what provides information about treatment effect heterogeneity. More specifically, we obtain the BLP parameters by estimating the following relationship in the main sample, using weighted OLS:

$$Y_i = \hat{\alpha}' X_{1,i} + \hat{\beta}_1(D_i - p(Z_i)) + \hat{\beta}_2(D_i - p(Z_i))(S_i - \mathbb{E}_N S_i) + \hat{\varepsilon}_i \qquad (2)$$

where $S(Z)$ is written as S for simplicity, $\mathbb{E}_N[w(Z_i)\hat{\varepsilon}_i X_i] = 0$ with $w(Z_i) = \{p(Z_i)(1 - p(Z_i))\}^{-1}$. Further, $X_{i,1}$ is constructed as $X_i = [X'_{1,i}, (D_i - p(Z_i), (D_i - p(Z_i)(S_i - \mathbb{E}_N S_i)]'$, and $X_{1,i}$ includes a constant, $B(Z_i)$, and $S(Z_i)$. In the above regression, the estimated β_1 is the ATE, and β_2 is best linear predictor of the existing heterogeneity. If what we estimate in the auxiliary sample ($S(Z)$) is a perfect proxy for the true heterogeneity, $s_0(Z)$, then $\beta_2 = 1$. If there is no heterogeneity, and the estimates from the ML are pure noise, then $\beta_2 = 0$.

We consider a range of possible baseline variables that could plausibly be correlated with outcomes: pre-exposure to 1AF program, household size, baseline fertilizer use, asset score, father above secondary education, self-reported credit access), as well as additional variables such as rainfall and temperature by growing season phase (pre-planting, immediate post-planting, and post-top-dressing). The results are qualitatively similar over many combinations and subsets of these variables. We present results generated using Neural Nets, but find similarly low levels of heterogeneity using a host of ML methods, including Lasso, Ridge, Elastic Net, Boosting, and Random Forest. In every instance, we have low ability to predict heterogeneity; the confidence intervals on $\beta_2 = 0$ are centered at zero, but are also imprecisely estimated. We do not report the results here, but they are available from the authors. This does not tell us with certainty that there exists no heterogeneity in treatment effects, but it does inform us that the vector of covariates included has no power to predict treatment effect heterogeneity.

We additionally 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. This suggests 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. Results of both exercises are available from the authors upon request.

G G – 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 20: 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		