Information, credit, and inputs: the impacts and mechanisms of a program to raise smallholder productivity*

Joshua W. Deutschmann† Maya Duru ‡ Kim Siegal § Emilia Tjernström†

February 23, 2019

Preliminary work; please do not cite

Abstract

Raising smallholder agricultural productivity has the potential to boost GDP and to reduce rural poverty, but the evidence on how to best achieve productivity gains remains mixed. Technologies exist that can increase smallholder yields and profits, but researchers and practitioners grapple with how to induce technology adoption in this population. This task is particularly challenging and pressing in sub-Saharan Africa, where agricultural yields lag all other regions and where poverty is often concentrated in rural areas. This paper presents randomized evidence of a rare scaled-up success story: One Acre Fund's (1AF) small farmer program. Much like anti-poverty "graduation" programs, 1AF's program is designed around the notion that farmers face multiple constraints simultaneously. Participating farmers receive input loans, crop insurance, and training about improved farming practices. Analyzing data from a pre-registered randomized control trial, we show that participation in 1AF's program causes statistically and economically significant increases in yields and profits. We find evidence that relaxing information constraints alone unlikely explains the program's success. Using various approaches to heterogeneity analysis, we find suggestive evidence that more disadvantaged farmers benefit more from the program and that program impacts increase over time. Both results are in line with the literature on graduation programs, which typically target the most disadvantaged, provide them with a bundle of different types of support, and continue to work with them as they climb out of poverty.

JEL codes: O12, O13, Q12

*Comments welcome: jdeutschmann@wisc.edu. We thank Brad Barham, Laura Schechter, Andrew Simons, Christopher Udry, William Masters, UW AAE Development Lab members, as well as participants at the IDEAS Summer School in Development and NEUDC for helpful comments and suggestions. All remaining errors are our own.

†University of Wisconsin-Madison

‡J-PAL

§One Acre Fund & George Washington University

1 Introduction

Raising smallholder productivity remains a key development challenge, and it has proven particularly intractable in sub-Saharan Africa. In 2014, US agricultural productivity exceeded African productivity by 473% (Magruder, 2018), and rural poverty remains much higher than urban levels (Beegle et al., 2016). The potential economic gains are significant: Gollin et al. (2018) estimate that a 10 percentage point increase in adoption of high-yielding crop varieties is associated with a 10-15 percentage point increase in GDP per capita. Furthermore, GDP growth originating in agriculture benefits the poor substantially more than growth originating in other sectors (Ligon and Sadoulet, 2008).

Agricultural policy experiments such as extension programs, fertilizer subsidies, and information interventions—tested by governments, NGOs, and researchers alike—have yielded substantial evidence on what fails and a limited catalogue of successes. Is this limited progress a sign that African agriculture is a lost cause or do existing programs lack something crucial? This paper uses a pre-registered, randomized control trial to examine the One Acre Fund (1AF) small farmer program, which succeeds in boosting smallholder productivity. By providing farmers with training, high-quality inputs on credit, and insurance, the program increases per-farmer yields by 32% and profits by 23% in our preferred specification.

The 1AF program's focus on credit, insurance, inputs, and information is no accident. The economics literature has accumulated substantial evidence that failures in any of these domains hinder farmers' ability or willingness to adopt improved agricultural technologies. Feder et al. (1985) emphasize three of these constraints in their review of the technology adoption literature at the time (credit, risk, and information). Magruder (2018) focuses on the same three constraints as key barriers to adoption in a more recent review of the experimental evidence on technology adoption. The focus on providing high-quality inputs also has empirical support: a growing literature establishes that fertilizer available on local markets in the region often falls short of quality standards (Bold et al., 2017; Tjernström et al., 2018).

Input loan interventions on their own have seen relatively low levels of takeup, and have had limited impacts on crop choices and productivity (Giné and Yang, 2009). Ashraf et al. (2009) examine a program that combines credit with increased access to export markets and find that bundling the two increased program takeup but failed to increase farmer incomes relative to farmers who were only offered the export program. Despite facing clear liquidity constraints, few farmers took up a voucher program for fertilizer and improved seeds in Mozambique (Carter et al., 2013). However, the farmers who did experiment with fertilizer in response to the vouchers continued to use the technology even after subsidies expired, suggesting a role for learning about optimal fertilizer use (Carter et al., 2014).

Theory clearly suggests that uninsured risk reduces input use, which in turn implies that crop insurance should increase input use. Empirical tests of this hypothesis show that when farmers receive insurance, they indeed increase their input use and shift towards higher-yielding but riskier crop varieties (Karlan et al., 2011; Mobarak and Rosenzweig, 2012; Elabed and Carter, 2014). Some studies suggest that insurance may be a more important driver of agricultural investment than credit (Karlan et al., 2014). Despite the promise that insurance holds, a large number of index insurance trials in a wide variety of contexts have been faced with very low take-up. We are unable to directly investigate the importance of the insurance component of 1AF's program, but will keep in mind the fact that farmers have access to crop and funeral insurance in our discussion of mechanisms below.[1]

In terms of information, farmers face a stochastic environment that makes it difficult to infer what optimal practices and technologies are. Even beyond the information needed to make adoption decisions about new technologies, farmers may persist in using sub-optimal agricultural practices because the noisiness of agricultural yields hampers their ability to learn. Providing detailed information about specific ways to improve yields can effectively shift farmer practices (Hanna et al., 2014). However, even if farmers do adopt improved practices, they may make mistakes that render adoption unprofitable (Duflo et al., 2008). Furthermore, farmers tend to learn and extrapolate from both their own experiences (Maertens et al., 2017) and those of their neighbors (Conley and Udry, 2010). This can help diffuse information through networks, but it may also perpetuate flawed information, especially when outcomes are highly stochastic and dependent on unobserved factors.

In addition to uncovering large average treatment effects (ATEs) from participation in 1AF's program, we also try to move beyond the ATEs to examine whether the results suggest that farmers face multiple simultaneous constraints. A large theoretical literature on poverty traps, and a growing

[1] Note that the insurance provided to 1AF farmers covers the cost of 1AF-provided inputs, but does not fully insure farmers against a failed harvest.

empirical body of evidence shows that poor households may need bundled interventions in order to move out of poverty (Bandiera et al., 2017; Banerjee et al., 2015). If we believe that farmers face multiple simultaneous constraints, offering an intervention that only relaxes credit constraints, or only provides information, may not be sufficient to raise yields and profits in a significant or economically meaningful way.

A reader could of course question whether this bundled approach is the most cost-effective one. Could a simpler program work as well at lower cost? Note first of all that the common criticism against bundled programs is typically aimed at donor-funded programs. The 1AF model, in contrast, is largely farmer-funded. Judging by the rates of re-enrollment in the program and by the high rates of loan repayments, farmers experience a substantial return on their investments.

To provide further support for a bundled program in this context, we follow the approach used in Dillon and Barrett (2017) to examine the relationship between household size and production and find evidence that suggests the presence of multiple input or output market failures. We argue these market failures boost the case for programs that bundle multiple interventions. Further, a bundled program that addresses multiple constraints may be effective across a greater number of contexts if the nature and severity of market failures varies enough across space. While opponents may contend that more complex programs are costlier to implement than more targeted interventions, this argument ignores the potential cost of tailoring targeted programs to suit new contexts. Determining what the binding constraints are in many small, local markets, likely requires substantial market research and could in many cases be prohibitively costly.

We analyze the question of multiple constraints in several ways. While our RCT did not randomly vary the program components (as in Banerjee et al. 2015), we believe this secondary analysis provides additional evidence that farmers face multiple constraints simultaneously, making bundled interventions necessary. First, we consider a sample of farmers with prior (non-randomized) access to the program. We observe that these farmers continue to use of many of the improved agricultural practices that the program teaches, but that these practices do not seem to translate into statistically significant yield gains. We interpret this as suggestive evidence that relaxing information constraints may not be sufficient on its own. Second, we conduct several types of heterogeneity analysis. We examine standard interactions with baseline characteristics, as well as new machine learning techniques, and find that farmers who use fewer improved inputs at baseline benefit more from the program. This suggests that the program particularly benefits more disadvantaged farmers, who are also more likely to face multiple constraints. Third, a comparison of farmers' fertilizer use on program plots to that on non-enrolled plots corroborates the notion that farmers are credit constrained. Finally, we examine farmers' re-enrollment decisions and the amount of land that they enroll and find that farmers who enroll in consecutive years tend to increase the amount of land that they enroll. This is consistent with past participation easing constraints, which then allows the farmer to increase their participation along the intensive margin. It is also consistent with farmers learning about the returns to the program and increasing their enrollment.

The rest of the paper is organized as follows. Section 2 describes the context of our study, the experimental design, and the different samples that we employ in our analysis. This section also familiarizes the reader with our sample by presenting summary statistics and balance checks and by comparing our sample to Kenyan farm survey that represents a larger section of the country. Section 3 describes our empirical analysis—both the pre-registered RCT analysis and the more exploratory investigations of mechanisms. Section 4 presents our primary empirical analysis of the effects of program participation on maize yields and profits, and discusses the improved practices that treated farmers adopt. Section 5 then investigates why pre-exposed farmers do not seem to be able to obtain the higher yields that we see among treated farmers, despite using many of the improved practices. We build on this examination to analyze whether information and credit seem to be binding constraints. Section 6 conducts a host of robustness checks of our main results, and Section **??** concludes.

2 Context and data

This study focuses on Kenya, where forty percent of the population lives in poverty, and a large share of the poor engage in smallholder farming as their primary occupation. The agriculture sector contributes 51 percent to the country's GDP (25% indirectly) and is dominated by small-scale producers (Government of Kenya, 2010). Despite the importance of the agricultural sector to the economy, most smallholder farmers are not running successful micro-enterprises. Households in Kenya typically derive their income from the production of a variety of crops, often combined with a range of off-farm activities (Sheahan et al., 2013). Most smallholder farmers produce less food than they need to feed their families, and are net maize buyers (Kirimi et al., 2011).

We further focus on maize—the main staple crop in Kenya—as the crop is crucial both to the economy and from a food security perspective. 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 maize varieties and inorganic fertilizer at higher rates than other countries in the region, maize yields remain low. Increasing yields and profits is crucial if we want the agricultural sector to act not simply as a means of subsistence, but as a pathway out of poverty.

1AF has operated in western Kenya for more than ten years, and served over 200,000 farmers in 2016. Their core "market bundle" addresses several of the key constraints discussed in Section 1: participating farmers receive group-liability loans for improved seeds and high-quality fertilizer, weekly training on modern agricultural techniques, crop and funeral index-based insurance, and market facilitation support to help farmers sell their products for higher prices.

Farmers choose the amount of land to enroll, and 1AF provides the agricultural inputs as a function of the amount of land enrolled. The group liability loans are given to self-selected farmer groups. The loan terms are quite flexible, allowing farmers to repay in any amount at any time during the growing season (but they must complete repayment in full by the end of the harvest). Historically, repayment rates have been over ninety-seven percent. Field officers conduct the weekly trainings with farmer groups; a field officer interacts on average with around 200 farmers.

2.1 Experimental design

The experiment was carried out in a cluster-randomized design in the Teso region of Kenya. As described above, farmers self-select into farmer groups of around 10-12 farmers. A cluster is then made up of 2-4 of these joint-liability farmer groups. Participants were recruited following standard 1AF protocol. Typically, once a farmer indicates interest in signing up and satisfies the basic criteria, he/she pays a small program participation deposit of approximately $5 USD.[2] For the purposes of the study, farmers also had to consent to take part. Shortly after contract signing, 1AF informed farmers about the study, that their participation would be voluntary, and further provided them with informed consent documents. 1AF informed farmers that half of them would be randomly

[2]To join 1AF, a farmer must satisfy the following restrictions: have a phone number and national identification; agree to repay their loan; and complete pre-payment of 500 Kenyan Shillings (Ksh, equivalent to approximately $5 USD).

assigned to treatment, while the other half would receive an alternative compensation package (household goods and a discount for 1AF participation the following season amounting to roughly 20% of the typical program cost). Randomization was conducted by public lottery, and all farmers enrolled in the study also received 10,000 Ksh in funeral insurance coverage, regardless of treatment status.[3]

The study is complicated by the fact that it took place in a region where 1AF has already operated for several years. The specific villages selected for study inclusion had never been offered the 1AF program, but neighboring villages had previously been offered the program. Thus, a substantial portion of farmers who expressed interest in participating had previously participated in the 1AF program (by "commuting" to the neighboring villages to participate). 1AF typically targets smallholder farmers living in areas with 100-500 people per km^2, who rely mostly on agricultural income, but who have limited access to farm inputs. Therefore, when 1AF decided to conduct an RCT to evaluate their work in Kenya, choosing completely new regions would have forced them to expand into regions that are quite unrepresentative of the typical program. For example, they would have had to enroll farmers in pre-dominantly tea-growing areas, or located in substantially different agro-ecological zones. Teso district in particular was chosen as a suitable region, being representative of 1AF program areas, while still having many farmers in the district who had not been previously exposed to 1AF.

Because of the "contamination" of some of the sample, we define two different samples for the analysis: the "primary sample," which consists of treated and control farmers who had never previously participated, and the "full sample," which includes the pre-exposed farmers. In some of our analyses, we also separately consider only the pre-exposed sample. Tables and graphs clearly note which sample we are using. It is worth keeping in mind that both of the samples present some challenges and some advantages. On the one hand, if the pre-exposed control farmers continue to benefit from their prior program involvement even after quitting the program (by, for example, continuing to use the new practices that they've learned), then treatment effects estimated with the full sample would likely result in a downward-biased impact estimate. On the other hand, we can examine the pre-exposed control farmers to see whether they achieve similar yields to the

[3]For the study, farmers also had to fulfill two additional requirements: they had to cultivate maize, and they had to be able to cultivate at least a quarter of an acre of maize.

newly-enrolled treated, and we can compare pre-exposed farmers in the treatment group to see whether effects accumulate with multiple years of treatment.

Furthermore, the farmers who had never been exposed may be different than those who self-selected into the program in earlier years. If selection into the program is positively correlated with potential returns to the program (i.e., higher-return farmers opt in earlier), we would expect the impacts on "hold-out" farmers to be a lower bound of the true impacts; if timing of enrollment is uncorrelated with returns to program participation, then we will obtain unbiased estimates of the program.

2.2 Data collection

The data collection for this experiment was directly managed by 1AF. To increase confidence in the process, they contracted with the International Initiative for Impact Evaluation (3ie) to help design and review all parts of the trial—including the experimental design, the field protocols, sampling and randomization, as well as the data collection instruments (Dubey and Yegbemey, 2017). 3ie concluded that the randomization was conducted successfully, and noted that 1AF staff showed high levels of professionalism in conducting the randomization.[4]

Readers may also worry about the potential for enumerator effects since 1AF collected the data from farmers. This arrangement was agreed upon for practical purposes as 1AF has dedicated and trained staff on the ground to collect the large amounts of household and physical harvest data that were required for this study. Several factors should help inspire confidence in the data collection: first, 1AF used standard best-practice protocols to ensure data quality, such as back-checks (resurveying 10% of respondents and comparing with original results) and in-field supervision. Second, 1AF contracted an independent survey firm, Intermedia Development Consultants (iDC), to carry out a three-step audit of the data collection[5]. The audit team participated in group meetings held by the 1AF monitoring team, 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. They

[4] A letter from 3ie attesting to their review and approval of the pre-analysis plan can be found in Appendix F.

[5] The iDC audit report is available from the authors upon request; the overall conclusion of the audit stated that "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."

observed the work of supervisors on 11 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. Finally, note that while social desirability bias could be a concern in some contexts if farmers or enumerators want to show the program in a good light. In our case, the main dependent variable was physically weighed, so should be an objective measure.

Two of the authors on this paper were brought in to review the pre-analysis plan (PAP) prior to follow-up data collection and to conduct the final analysis according to the PAP. The team carried out all the variable construction and analysis from scratch, i.e. without referring to 1AF's existing analytical code.[6] This paper extends the basic impact evaluation analysis from the report in order to investigate impact mechanisms; Table A.1 in Appendix A clearly delineates which analyses were pre-registered and which fall outside of the purview of the PAP.

Baseline data collection occurred in November and December of 2016—after program enrollment but before treatment assignment. The public lottery, which assigned clusters of farmer groups to treatment, took place in January 2017. Enumerator teams rolled out input use surveys after the planting of the main season in 2017, from April through June. These data provide detailed information on the extent of farmer compliance with treatment, such as whether farmers spaced their plants correctly, and applied the correct fertilizer dosage. We use these data as indicators of the extent to which farmers are actually learning and changing their behaviors as a result of the 1AF training.

The main outcome variables come from a harvest data collection effort that 1AF carried out during the main maize and beans harvest period (May to October, 2017). The procedure for collecting yield data is the following: for treatment farmers, enumerators randomly selected one maize plot that was enrolled in the program ("enrolled plot") and one maize plot that was not enrolled ("non-enrolled plot"). For control farmers, enumerators randomly selected a random maize plot. Enumerators then collected wet and dry harvest weights from two randomly placed 8 x 10 boxes for each selected plot. These estimates are applied to all cultivated land to determine per-farmer yields (for treatment farmers, we compute a weighted average of 1AF and non-1AF cultivated land size). We expect this method to produce substantially better yield estimates than

[6] The final report can be found here: Deutschmann and Tjernström (2018) [link to report]

the self-reports commonly reported by economics studies. Furthermore, we measure land size based on GPS readings of each field (we can also compare this to self-reported land size; more on this in Section 6).

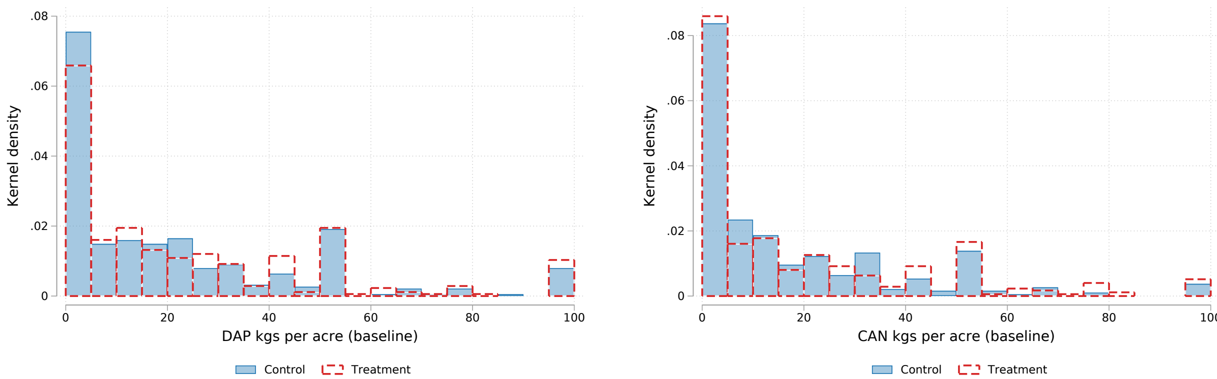
2.3 Summary statistics and baseline balance

Table 1: Baseline balance across treatment and control groups

	(1) Control		(2) Treatment		Difference
Variable	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	(1)-(2)
Married	1062	0.878	1075	0.880	-0.002
	[60]	(0.012)	[60]	(0.011)	
Father, 2ary school (0/1)	1062	0.387	1075	0.427	-0.040
	[60]	(0.020)	[60]	(0.024)	
Farm labor >50% income	1062	0.779	1075	0.779	0.000
	[60]	(0.020)	[60]	(0.019)	
Used ag tech 2016	1062	0.782	1075	0.802	-0.019
	[60]	(0.025)	[60]	(0.017)	
Prev. 1AF knowledge	1062	0.467	1075	0.525	-0.058
	[60]	(0.033)	[60]	(0.026)	
Intercropped 2016	1062	0.477	1075	0.475	0.002
	[60]	(0.027)	[60]	(0.031)	
Credit access 2016	1062	0.711	1075	0.725	-0.014
	[60]	(0.022)	[60]	(0.020)	
Household size	1062	6.655	1075	6.822	-0.167*
	[60]	(0.095)	[60]	(0.099)	
Maize acres, 2016	1062	1.006	1075	1.025	-0.019
	[60]	(0.041)	[60]	(0.035)	
Maize yield/acre, 2016	1062	534.101	1075	580.891	-46.790*
	[60]	(28.965)	[60]	(28.112)	
F-test of joint significance (F-stat)					1.062
Number of observations					2137

Notes: Field office fixed effects are included in all estimation regressions. Standard errors clustered at farmer group cluster level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Who are our sample farmers? Table 1 presents baseline means of key control variables in the full sample, as well as balance tests across treatment and control samples. The majority of the sample is married, and less than half of male household heads have completed secondary school. The majority of households (just under 80%) earned more than 50% of their income from farming in the last year, corroborating that farming is a key activity in this population. The average farmer plants one acre of maize and gets roughly half a ton (500 kgs) of maize per acre.



(a) (b)

Figure 1: Fertilizer use at baseline

Many households use some type of agricultural technology at baseline (roughly 77%), but this basic variable does not capture the intensity of agricultural technology use. Figures 1a and 1b show the baseline fertilizer intensity of two key inorganic fertilizers, DAP (diammonium phosphate) and CAN (calcium ammonium nitrate). These two graphs clearly show that while many farmers used some fertilizer at baseline, most farmers applied them at low rates.

About half the sample answer that they have knowledge of 1AF planting practices ("Previous 1AF knowledge"). This may sound high, but is primarily driven by the pre-exposed farmers, and the levels are much lower in the primary sample: six percent in the control group and fourteen percent in the treatment group (details in Appendix B).[7]

Only two variables are statistically different across the two groups in the full sample: household size and self-reported per-acre yields.[8] All of our results are robust to controlling for past 1AF knowledge, as well as considering only the sample with no baseline 1AF knowledge. Baseline maize yields are by chance almost nine percent larger in the treatment group, a difference that is statistically significant at the 10% level. While we are confident that the randomization worked well, we control for baseline maize yields and the other variables in this table in all regressions

[7]This difference is statistically significant in the primary-only sample.

[8]An F-test of joint orthogonality of the variables in Table 1 is not significant in the full sample, but it is significant at the 10% level when we examine the primary sample on its own. The separate balance tables for the primary sample and the pre-exposed sample are in Appendix B.

(as specified in our pre-analysis plan). None of the results are sensitive to the inclusion of these additional controls.

2.3.1 Pre-exposed vs. new participants

Table 2 compares the farmers across the pre-exposed and new participant samples. It is clear that the two samples differ substantially. Pre-exposed farmers, who previously self-selected into the program, are more likely to have an educated father in the household, more likely to earn at least 50% of household income from farm labor, more likely to report using improved seeds or fertilizer at baseline, more likely to report baseline knowledge of 1AF practices, less likely at baseline to intercrop beans with maize, more likely to report having access to credit, reported using more land for maize at baseline, have bigger households, and reported higher maize yields per acre at baseline.

Several of these differences (education, land endowments, yields, etc.) suggest that farmers who enrolled earlier may be better farmers—although several of these variables were likely directly affected by program participation. Because of this, in most cases we prefer to present results using the primary sample, despite the loss in power afforded by the smaller sample size. However, our results are robust to using the full sample of both primary and pre-exposed farmers.

2.4 External validity

In any study that uses a non-representative sample, we may worry that results would fail to translate to other regions, let alone countries. In our study, since the primary sample consists of farmers who were late to join the 1AF program, we might additionally wonder how different they are from farmers who were early-joiners since we have already seen that they are different from the pre-exposed farmers. To compare our sample to a more representative sample, we use data from a household survey collected by Tegemeo Institute of Agricultural Policy and Development under the Tegemeo Agricultural Policy Research and Analysis (TAPRA) program. We use the latest round of data that is publicly available, which was collected in 2010. While this is seven years earlier than our data collection, average maize yields have not changed dramatically over that period. The data collection was designed to represent a diverse set of agro-ecological maize-growing zones across the country, making it particularly appropriate data set to compare our sample farmers to.

Figure 2 compares the distribution of acres planted to maize across the full Tegemeo sample

Table 2: Baseline balance across primary and pre-exposed samples

	(1) Primary Sample		(2) Pre-Exposed Sample		Difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)
Married	765	0.893	1372	0.871	0.022
		(0.011)		(0.009)	
Father, 2ary school (0/1)	765	0.362	1372	0.432	-0.070***
		(0.017)		(0.013)	
Farm labor >50% income	765	0.771	1372	0.783	-0.012**
		(0.015)		(0.011)	
Used ag tech 2016	765	0.635	1372	0.880	-0.244***
		(0.017)		(0.009)	
Prev. 1AF knowledge	765	0.101	1372	0.716	-0.616***
		(0.011)		(0.012)	
Intercropped 2016	765	0.515	1372	0.455	0.060***
		(0.018)		(0.013)	
Credit access 2016	765	0.707	1372	0.724	-0.017***
		(0.016)		(0.012)	
F-test of joint significance (F-stat)					150.862***
Number of observations					2137

Notes: Field office fixed effects are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

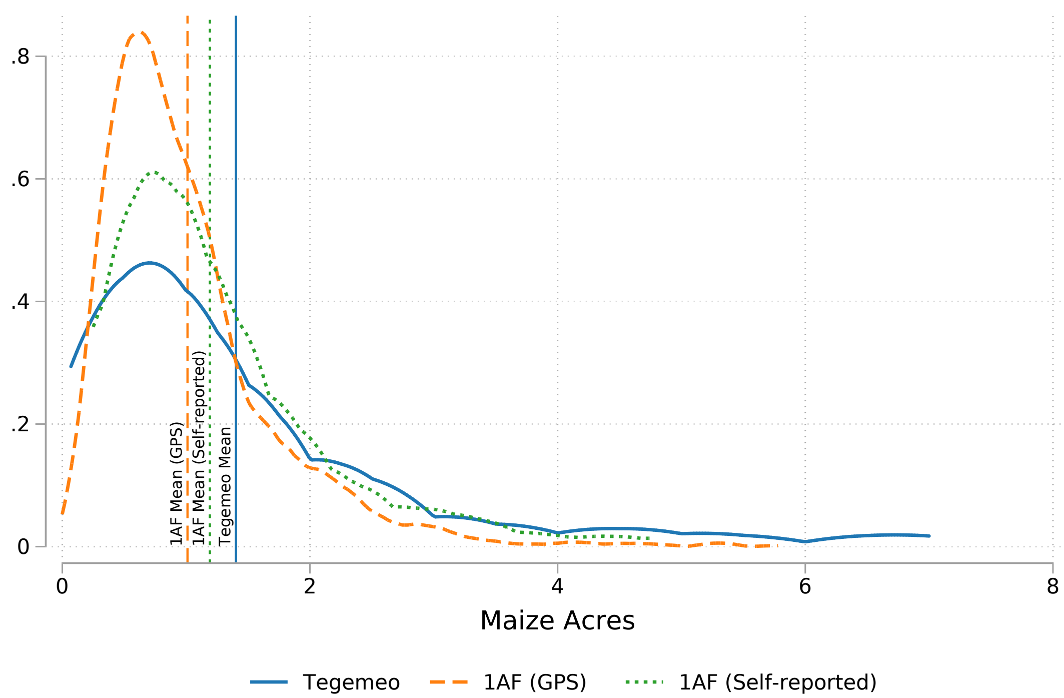


Figure 2: Acres planted to maize in our data and Tegemeo data

and our sample from western Kenya. We include two versions of maize acreage from our data: self-reported acres elicited from farmers at baseline, and our GPS-recorded acreage collected at follow-up. The mean acreage in the Tegemeo data is marked by the solid blue horizontal line, and is close to 1.5 acres. Our sample farmers have smaller fields, but there is also a substantial difference between the self-reported data (dashed orange line) and that measured using GPS (dotted green line). This suggests that at least along the dimension of land endowment, our sample farmers are on average slightly less well off compared to a more geographically diverse sample, but we can also see that some of the differences stem from differences in measurement.

Figure 3 compares the per-acre maize yields that farmers in the different samples obtain. Here, the mean per-acre yields in our primary sample at baseline and in the Tegemeo sample are very similar, while the pre-exposed farmers obtained substantially higher yields at baseline. Note that some of the pre-exposed farmers were actually enrolled in the 1AF program during our baseline farming season, which likely boosted their yields.

In sum, at least on these basic observable characteristics, it does not appear that our sample differs dramatically from other smallholders in Kenya. They may still differ on other key variables and on unobservables, but we hope that this comparison puts our sample a bit more in context and helps the reader interpret our results.

3 Empirical strategy

This section discusses our analytical approach, both for the primary analysis (as pre-specified in the PAP) and for the additional analyses that try to dig into impact heterogeneity and mechanisms.

3.1 Pre-registered main analysis

The main analysis focuses on three outcomes of interest: per-acre maize yields, per-farmer maize yields, and profit (in USD). Maize yields use dry weight measurements from randomly-selected areas on farmers' plots. For treated farmers, the per-farmer yield outcome accounts for yields on enrolled and non-enrolled land. Profits account for program costs, input costs, labor costs, and average local market prices over the off-season. Revenues for profit calculations account for both maize and any inter-cropped bean yields, since around half of the sample farmers intercrop (it is difficult to

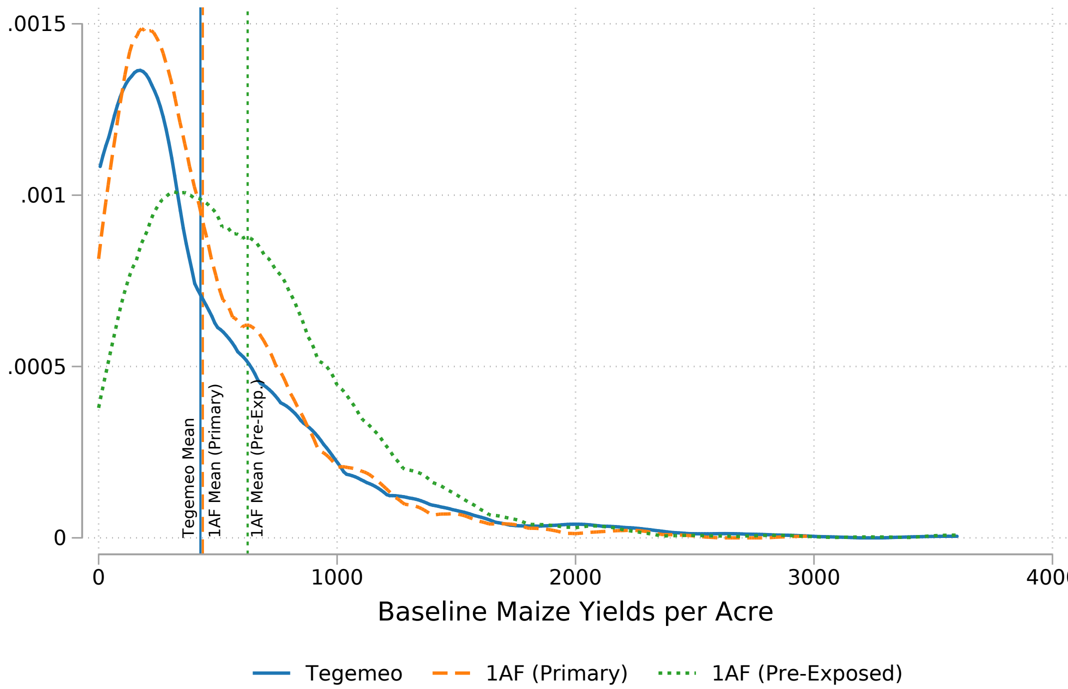


Figure 3: Baseline yields in our data (primary sample and pre-exposed) and Tegemeo data

separate labor or fertilizer costs across crops within the same plot). See Appendix E for a more detailed explanation of how the outcome variables are constructed.

For each outcome, we estimate the following regression:

$$Y_{is} = \alpha + \beta T_{is} + \delta X_{is} + \gamma_s + \epsilon_{is} \tag{1}$$

where T_{is} is the treatment dummy for individual i in field officer area s, X_{is} includes the list of pre-specified controls, γ_s is a field office area fixed effect, and ϵ_{is} is clustered at the cluster (treatment assignment) level.

To ensure correct inference given our multiple hypotheses, we follow Jones et al. (2018) and implement Westfall and Young (1993) free step-down re-sampling to control for family-wise error rate. We treat each regression table as a family of hypotheses. We present adjusted p-values based on this method at the bottom of each table in Section 4. The method does not change the significance level of any of the treatment coefficients.

3.2 Heterogeneity analysis

For the first part of the heterogeneity analysis, we estimate a regression in which we interact the treatment dummy with binary indicators of various baseline characteristics, as in the following:

$$Y_{is} = \alpha + \beta_1 T_{is} + \beta_2 K_{is} + \beta_3 (T_{is} \times K_{is}) + \delta X_{is} + \gamma_s + \epsilon_{is} \tag{2}$$

where K_{is} is the key variable of interest for heterogeneity analysis (for individual i in field officer area s). Most of our analysis will show the point estimate and 95-percent confidence interval of the average marginal effect of treatment for farmers with farmers with $K_{is} = 1$ and $K_{is} = 0$.

For a second set of heterogeneity results, we try to understand more broadly whether the ATEs hide substantial treatment effect heterogeneity—across a large number of covariates—and whether observable variables are informative about which farmers benefited the most or the least from the program. We do this following the recently developed machine learning methods in Chernozhukov et al. (2018) to conduct inference on several features of heterogeneous treatment effects. The approach is agnostic about specific machine learning estimators, and instead focuses on generating valid inference about *features* of the heterogeneous effects—accounting for the added uncertainty introduced by sample splitting. In this analysis, we estimate two key features: Sorted Group Average Treatment Effects (GATEs), which bins observations by their predicted treatment effect, and examine the estimated treatment effects across five such groups. Second, Classification Analysis reveals how covariates of interest differ between the units that are grouped into the lowest and the highest groups from the GATEs analysis, to examine whether they differ systematically. We additionally report the Conditional Average Treatment Effect (CATE) estimates and Heterogeneity Loading Parameters (HET) from the two “best” estimators.

4 The impacts of 1AF participation on yields and profits

This section presents our main results on program impact, in which we evaluate the impact of the 1AF market bundle on yields and profits. Table 3 shows result using our preferred specification and the smaller (not pre-exposed) sample. Column (1) shows the impact of the treatment on per-acre maize yields, column (2) shows effects on per-farmer yields, and column (3) shows impacts on profits.

These correspond to a 28% increase in per-acre yields, a 32% increase in per-farmer yields, and a 23% increase in profits.

The magnitude of the profit results is particularly remarkable since even programs that affect yields often fail to increase profits. Further, treatment farmers spent a mean of $126 USD (median $108) on inputs, labor, and program participation costs, compared to a mean of $87 (median $73) for control farmers. Using this primary sample estimate of the treatment effect on profits, this suggests a 63% return on average for farmers who participated. Similarly, if we compare the estimated coefficient with self-reported income at baseline (mean $751, median $481), this increase in profit is far from trivial.

While we report our preferred specification here (including covariates, site fixed effects, and using the primary sample), we additionally conduct a number robustness checks in Section 6, hopefully convincing even skeptical readers that our results do not hinge upon these model assumptions. We demonstrate that these results are robust to different sample definitions and to the exclusion of the covariates and of site fixed effects. Also reported in Section 6) are results using randomization inference, inspired by Young (2017).

We also examine whether the treatment affected farmers' practices. Table 4 shows results examining the effect of 1AF participation on various improved farming practices. Column (1) looks at the treatment impact on farmer likelihood of using a commercial improved seed. The program does not drive a change in this practice, despite commercial seed being one of the main components of the 1AF bundle. This could be due to high baseline use of commercial seed by farmers in our sample. Column (2) shows treatment impacts on the likelihood of intercropping with beans, a practice that 1AF recommends. The lack of impact here could again be limited by those farmers who were already adopting that practice at baseline.

Columns (3) - (6) examine fertilizer use. DAP (diammonium phosphate) is a fertilizer that should be applied at the time of planting. It is the fertilizer type that farmers are most likely to be familiar with. However, anecdotally some farmers incorrectly apply it also later in the season, which tends to be ineffective. CAN (calcium ammonium nitrate) on the other hand is less commonly used in this area. It should not be applied at planting, but should instead be applied when the maize plant is about knee-high. Here, we do not examine application rates, but rather the timing of fertilizer application. In columns (3) and (4), we can see that the treatment increases the likelihood

Table 3: Primary outcomes, primary sample

	(1) Maize PA	(2) Maize Per Farmer	(3) Profit
1AF participant	316.088***	345.074***	79.834***
	(36.756)	(84.088)	(29.996)
Married	107.845*	-41.144	-19.811
	(58.243)	(152.252)	(55.475)
Father, 2ary school (0/1)	-6.988	249.461***	85.171**
	(45.432)	(94.245)	(34.894)
Farm labor >50% income	18.624	-10.724	1.530
	(49.214)	(84.821)	(31.330)
Used ag tech 2016	-89.555**	-100.434	-49.228
	(42.514)	(86.068)	(30.826)
Prev. 1AF knowledge	-11.078	114.447	22.103
	(65.881)	(166.821)	(61.844)
Intercropped 2016	-22.063	-138.885	-45.234
	(43.543)	(94.922)	(34.524)
Credit access 2016	-86.411**	-65.529	-28.493
	(34.569)	(84.676)	(30.171)
Household size	8.502	35.935**	13.854**
	(7.671)	(17.200)	(6.035)
FAW Incidence	-5.972	-1.091	0.668
	(3.655)	(8.499)	(3.177)
Maize acres, 2016	58.901**	413.300***	126.476***
	(28.413)	(79.716)	(28.411)
Maize yield/acre, 2016	0.174***	0.461***	0.149***
	(0.047)	(0.108)	(0.039)
Observations	765	637	637
R^2	0.185	0.282	0.236
MHT p-value for 1AF	0.000	0.000	0.009
Control Mean Dep. Var	1118.684	1083.774	339.788

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions which include field office area fixed effects. Columns (2) and (3) exclude observations missing non-program yields (see Appendix C). MHT p-value uses Jones et al. (2018) implementation of Westfall and Young (1993) correction for multiple hypothesis testing.

* $p < .1$, ** $p < .05$, *** $p < .01$

that a farmer applies DAP fertilizer at planting, but has no effect on the at-planting application of CAN. This is encouraging, as it seems farmers are applying the fertilizer at the recommended time. The post-planting behavior in columns (5) and (6) also suggest that treated farmers are more likely to apply CAN and to do so in the recommended manner (post-planting). We will examine fertilizer rates in more detail in Section 5.

We can also examine several other recommended practices, including whether the farmer prepared her field with a plow, and whether the spacing of plants and rows within the field are according to the recommended agronomic practices. Here, we define farmers as doing it correctly if their average row- and plant-spacing practices are within 5 cm of the recommended method. We can see that the treatment substantially increases the likelihood of both correct row spacing and plant spacing. Since plant density tends to deviate from agronomic best-practice in most of sub-Saharan African smallholder agriculture, this improvement could have important implications for yields. Plant spacing is also a practice that should be less costly than some of the other recommended practices like applying purchased inputs like fertilizer and commercial seeds.

We additionally look at the question of improved practices as an index. We consider four key recommendations: use DAP at planting, use CAN after planting, correct row-spacing (within 5 cm of the recommended 75 cm apart), and correct plant-spacing (each plant within a row should be approximately 25 cm from the next). Here we can look at difference between treatment and control groups in terms of the number of practices that they adopt. Figure 4 shows, within the primary sample, that treated farmers indeed adopt more of the recommended practices than do control farmers, although control farmers did already tend to adopt some of the same practices.

We have thus established that the treatment affects the agricultural practices of treated farmers, and that the program significantly increases yields and profits. From a policy and implementation perspective, it seems logical to wonder to what extent the training and subsequent improvements in improved practices drives yield and profit impacts. Since fertilizer and seed are costly and risky investments, adoption of low-cost improved practices may be easier to diffuse than other technologies—if indeed they drive the results.

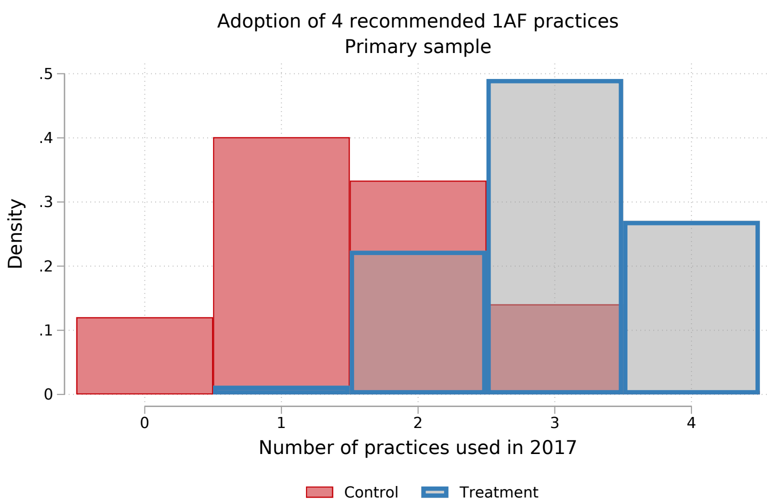


Figure 4: Improved practice adoption by treatment status

Table 4: Study year (2017) practices, primary sample

	(1) Imp. Seed	(2) Int. Beans	(3) CAN AP	(4) DAP AP	(5) CAN PP	(6) DAP PP	(7) Used Plow	(8) Row Spacing	(9) Plant Spacing
1AF participant	0.020	0.020	-0.016	0.583***	0.295***	-0.161***	0.047	0.321***	0.280***
	(0.019)	(0.031)	(0.009)	(0.032)	(0.025)	(0.026)	(0.034)	(0.041)	(0.032)
Married	0.002	-0.003	-0.008	-0.029	0.014	-0.018	-0.018	0.091*	0.127***
	(0.041)	(0.052)	(0.017)	(0.042)	(0.048)	(0.041)	(0.049)	(0.053)	(0.044)
Father, 2ary school (0/1)	0.005	0.009	0.014	-0.020	0.059**	-0.032	0.052*	0.070*	-0.056*
	(0.022)	(0.027)	(0.010)	(0.026)	(0.030)	(0.033)	(0.027)	(0.037)	(0.029)
Farm labor >50% income	0.015	0.038	-0.012	-0.021	0.004	0.037	-0.012	-0.053	-0.001
	(0.031)	(0.036)	(0.012)	(0.033)	(0.027)	(0.031)	(0.044)	(0.049)	(0.036)
Used ag tech 2016	0.551***	0.009	-0.000	0.056*	0.073**	-0.013	0.059*	-0.073*	0.044
	(0.045)	(0.032)	(0.010)	(0.030)	(0.035)	(0.028)	(0.036)	(0.039)	(0.037)
Prev. 1AF knowledge	-0.110***	-0.040	-0.007	-0.051	-0.043	-0.020	0.073*	0.061	-0.041
	(0.033)	(0.051)	(0.012)	(0.035)	(0.043)	(0.041)	(0.038)	(0.062)	(0.045)
Intercropped 2016	0.045*	0.124***	-0.006	0.003	-0.016	0.047*	0.020	-0.027	-0.056*
	(0.024)	(0.031)	(0.007)	(0.030)	(0.030)	(0.027)	(0.025)	(0.035)	(0.031)
Credit access 2016	-0.008	-0.001	0.017*	0.047*	-0.029	-0.062**	-0.035	-0.044	-0.085**
	(0.019)	(0.030)	(0.010)	(0.028)	(0.033)	(0.025)	(0.031)	(0.039)	(0.034)
Household size	0.003	0.000	-0.002	0.003	-0.003	-0.000	0.003	-0.011	0.003
	(0.005)	(0.006)	(0.002)	(0.006)	(0.006)	(0.005)	(0.005)	(0.007)	(0.007)
FAW Incidence	0.004*	-0.001	0.001	-0.009***	-0.003	0.002	0.003	0.002	0.004
	(0.002)	(0.003)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Maize acres, 2016	0.043***	-0.008	-0.004	-0.017	0.014	0.019	0.043**	0.012	-0.021
	(0.015)	(0.022)	(0.004)	(0.022)	(0.019)	(0.016)	(0.020)	(0.026)	(0.016)
Maize yield/acre, 2016	0.000***	-0.000	0.000	-0.000	0.000	0.000**	0.000*	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	765	762	765	765	765	765	765	765	765
R^2	0.516	0.343	0.037	0.456	0.190	0.231	0.282	0.142	0.200
Ctrl Mean Dep. Var	0.802	0.417	0.023	0.391	0.657	0.273	0.732	0.371	0.090

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions which include field office area fixed effects. Imp. Seed refers to use of commercial improved seed. AP stands for at planting, and PP stands for post planting. 1AF recommends using DAP at planting and CAN post planting. Row spacing and plant spacing are 1 if average spacing within 5 cm of recommended.

* $p < .1$, ** $p < .05$, *** $p < .01$

5 Do program participants face multiple constraints?

The 1AF bundle was designed to address several constraints that we often believe that farmers face. If farmers are stuck in a poverty trap, or for other reasons need a "big push" intervention, then a bundle of complementary interventions is likely required. This could then help explain why the 1AF program has been successful where other programs have largely failed. However, if instead we could place farmers onto higher trajectories by relaxing a single constraint, then the organization could potentially reduce costs.

5.1 Separability

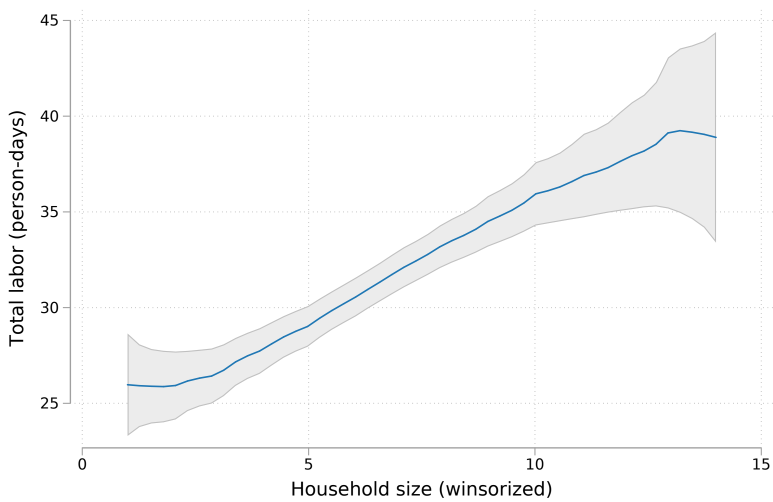


Figure 5: Number of person-days employed on the farm as function of household size

The literature on competitive markets in rural sub-Saharan Africa provides an eagle-eye view on why we should take seriously the notion of a bundled program as potentially necessary for success in this context. Agricultural factor markets in sub-Saharan Africa are often failing or incomplete, as evidenced by a growing literature that derives and tests the necessary conditions consistent with failing markets. For example, the reduced-form test in Benjamin (1992)[9] has been applied to sub-Saharan Africa (Udry, 1996; Dillon and Barrett, 2017) and typically resoundingly rejects the hypothesis that input and output markets are well-functioning.

Here, we replicate the standard test of a key prediction of the complete market hypothesis: the

[9]See also LaFave and Thomas (2016) for an extension using panel data.

separability between household production and consumption decisions in our context. The test is derived from the notion of *separability*, which comes from the agricultural household model and implies that that agricultural households in well-functioning input and output markets should be able to make production and consumption decisions separately. If separability holds, land and labor endowments should not influence production decisions as these decisions should come from profit-maximization on the farm. We find that key household demographic measures (household labor endowments) are highly correlated with the amount of farm labor that households employ in production (see Figure 5 and Appendix Tables G.1 - G.4).

Following Dillon and Barrett (2017), we measure labor in person-days used for land preparation, planting, and post-planting activities, accounting for child person-days at 50% of adult person-days. The clear positive relationship between household size and total labor use shown in Figure 5 constitutes suggestive evidence that separability does not hold for our sample farmers. Table G.1 presents a more formal test of this in the context of our primary sample. In particular, we see that land endowment (acres owned & cultivated) is positively correlated with total labor demand, as is household size (column (3)). This is similarly true if we consider the pre-exposed sample (Appendix Table G.2) and if we account for zeros in labor demand by instead using an inverse hyperbolic sine transformation (Appendix Tables G.3 and G.4). This suggests failures in at least two input or output markets, which may be an additional reason why the bundled program that we evaluate has succeeded where others have failed.

5.2 Information

A key component of the 1AF bundle is the training that farmers receive on optimal planting practices. As we saw in Section 4, the program does increase farmers' adherence with recommended agronomic practices, suggesting that there may have been some information gaps at the baseline. If farmers continue to apply these improved practices after they are no longer participants in the program, it suggests that they learned from the program that these practices are beneficial.

Table 5 shows suggestive evidence that pre-exposed farmers tend to use improved practices. The table only includes control farmers, and our coefficient of interest is that on the dummy for pre-exposure. Of course, since pre-exposure to 1AF's programming is non-random, it is entirely possible that farmers who selected into the program earlier were different than those who delayed

Table 5: Study year (2017) practices, control sample

	(1) Imp. Seed	(2) Int. Beans	(3) CAN AP	(4) DAP AP	(5) CAN PP	(6) DAP PP	(7) Used Plow	(8) Row Spacing	(9) Plant Spacing
Past 1AF participant	0.123***	-0.014	-0.011	0.089**	0.057*	-0.060**	0.073**	0.068*	0.019
	(0.023)	(0.027)	(0.009)	(0.036)	(0.032)	(0.028)	(0.032)	(0.037)	(0.023)
Married	0.080**	0.004	-0.020	-0.049	-0.008	0.010	-0.038	0.047	0.063*
	(0.036)	(0.045)	(0.015)	(0.039)	(0.040)	(0.042)	(0.032)	(0.042)	(0.036)
Father, 2ary school	0.007	0.000	0.010	0.030	0.075***	-0.014	0.049**	0.055	0.023
(0/1)	(0.023)	(0.027)	(0.008)	(0.028)	(0.025)	(0.025)	(0.019)	(0.039)	(0.020)
Household size	0.002	0.010*	-0.001	0.004	0.002	-0.006	0.012**	-0.008	0.003
	(0.004)	(0.006)	(0.001)	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.005)
FAW Incidence	0.004**	0.000	0.001	-0.001	-0.000	0.001	0.005***	0.002	-0.002
	(0.002)	(0.003)	(0.001)	(0.003)	(0.003)	(0.002)	(0.002)	(0.004)	(0.002)
Observations	1062	1056	1062	1062	1062	1062	1062	1062	1062
R^2	0.065	0.235	0.023	0.136	0.092	0.170	0.257	0.036	0.084
Ctrl Mean Dep. Var	0.878	0.491	0.015	0.489	0.724	0.258	0.763	0.408	0.122

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions which include field office area fixed effects. Comm. Seed refers to use of commercial (improved) seed. AP stands for at planting, and PP stands for post planting. 1AF recommends using DAP at planting and CAN post planting. Row spacing and plant spacing are 1 if average spacing within 5 cm of recommended.

* $p < .1$, ** $p < .05$, *** $p < .01$

enrollment. If early-adopters were better farmers, they may have been using some of the improved practices even in the absence of the program. In particular, the coefficient on past participant in column (2), which shows that pre-exposed farmers are more likely to plant a commercial seed than are non-pre-exposed farmers suggests that selection is likely a part of the story here.

With that in mind, we can note that the pre-exposed farmers are more likely to use DAP at planting (column 4), and correctly less likely to use DAP post-planting (column 6). They are also more prone to using a plow and to correctly space their maize (columns 8 and 9). While we cannot guarantee that the self-selected farmers did not already use these practices before enrolling, the fact that the adoption lines up quite well with the (randomized) treatment effect on practices suggests that at least some of these practices are likely persistent effects of the treatment. So with all these improved practices, do pre-exposed farmers also reap the higher yields associated with 1AF participation?

Table 6 investigates this by regressing our main outcome variables on a dummy for past 1AF participation, within the control sample. As we can see, even though we have seen that past participants employ several of the recommended practices, this does not appear to translate into yield or profit impacts. The point estimates are positive, but not statistically significant. To prod this a little further, we also examine whether treatment impacts vary by baseline improved practice adoption.

If practices matter, or drive the effects of the program, we might also expect farmers who already adopt improved practices to benefit less from the program than farmers who are less well-informed. Table 7 carries out an analysis with our primary sample to address this question. While the data on improved practices is less detailed in our baseline data than in the follow-up, we have a few ways of measuring improved practices. Column (1) of Table 7 interacts treatment with a dummy for whether or not the farmer already used an improved seed variety at baseline. This interaction term is negative and significantly different from zero. Columns (2) and (3) carry out a similar analysis for whether or not the farmer used DAP or CAN, and while the interaction terms are negative. they are not statistically significant.

Columns (4) and (5) explore instead an index indicating the number recommended practices that a farmer adopted at baseline. This index is different from the one discussed in Section 4 as we don't have information on aspects like row spacing at baseline. The first of the two columns

Table 6: Primary outcomes, control sample

	(1) Maize PA	(2) Maize Per Farmer	(3) Profit
Past 1AF participant	26.065	119.789	35.584
	(37.045)	(72.068)	(27.018)
Married	89.682*	155.821	49.716
	(49.243)	(103.261)	(37.192)
Father, 2ary school (0/1)	45.438	350.511***	115.558***
	(38.804)	(84.684)	(30.430)
Household size	3.047	55.615***	18.060***
	(6.172)	(14.750)	(5.215)
FAW Incidence	-5.514	-1.355	-0.334
	(3.551)	(6.233)	(2.281)
Observations	1062	1062	1062
R^2	0.072	0.124	0.109
MHT p-value for 1AF	0.484	0.102	0.193
New Participant Mean Dep. Var	1118.684	1083.774	339.788

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions which include field office area fixed effects. MHT p-value uses Jones et al. (2018) implementation of Westfall and Young (1993) correction for multiple hypothesis testing.

* $p < .1$, ** $p < .05$, *** $p < .01$

employs a dummy variable equal to one if the farmer adopted one or more recommended practices at baseline, and the second includes an indicator variable for each level of the index. Again, all the interaction terms are negative but not statistically different from zero.

Taken together, these results provide some evidence that farmers learn from the 1AF program— i.e., that information was indeed a constraint— and that they continue using improved practices to some extent, but that these improved practices do not translate into substantially improved yields. Table 7 suggests that the farmers who had access to improved seeds at baseline reap less benefits from the program. Foreshadowing the analysis in the next section, this perhaps hints that access—perhaps to improved inputs or to credit to purchase those inputs—play an important role.

5.3 Credit

The simplest test of whether credit matters may be to examine whether or not treatment impacts vary by baseline credit access. In other words, do farmers for whom some of the key constraints that the program targets were less binding at baseline appear to reap lower benefits from the treatment? For example, if the credit component of the program were the key to its success, we might imagine

Table 7: Yields per acre and baseline practice use, primary sample

	(1)	(2)	(3)	(4)	(5)
1AF participant	436.751***	365.451***	301.129***	537.770***	538.358***
	(89.966)	(62.914)	(58.231)	(179.062)	(179.670)
Used Imp. Seed=1	40.524				
	(58.145)				
Used DAP=1		50.366			
		(55.037)			
Used CAN=1			33.465		
			(58.714)		
Used any improved practices=1				165.172*	
				(96.409)	
Nb of improved practices=1					141.789
					(97.138)
Nb of improved practices=2					179.996*
					(103.832)
Nb of improved practices=3					164.647
					(102.778)
1AF participant × Used Imp. Seed=1	-164.130*				
	(98.727)				
1AF participant × Used DAP=1		-88.032			
		(77.744)			
1AF participant × Used CAN=1			10.733		
			(71.792)		
1AF participant × Used any improved practices=1				-246.820	
				(182.870)	
1AF participant × Nb of improved practices=1					-217.650
					(190.712)
1AF participant × Nb of improved practices=2					-245.842
					(191.111)
1AF participant × Nb of improved practices=3					-263.641
					(189.395)
Married	106.419*	111.097*	105.838*	106.068*	108.626*
	(61.789)	(62.055)	(61.429)	(62.043)	(62.583)
Father, 2ary school (0/1)	6.321	2.482	0.996	3.992	3.766
	(44.476)	(44.724)	(44.405)	(44.322)	(44.046)
Household size	11.843	10.855	11.057	11.483	11.636
	(7.363)	(7.150)	(7.163)	(7.259)	(7.440)
FAW Incidence	-5.080	-4.830	-4.911	-5.109	-5.107
	(3.760)	(3.658)	(3.672)	(3.649)	(3.632)
Observations	765	765	765	765	765
R^2	0.168	0.165	0.165	0.167	0.168

Standard errors in parentheses

Note: standard errors clustered at farmer group level. The outcome variable is maize yields per acre. Results in this table are from OLS regressions which include field office area fixed effects. Imp. Seed refers to use of commercial improved seed.

* $p < .1$, ** $p < .05$, *** $p < .01$

that farmers who already had access to credit at baseline may see less of a boost from participating.

5.3.1 Direct analysis of credit constraints

We have in our data two measures of credit access: a narrower definition that includes only respondents who reported having received a loan at baseline ("reported loan use"), and a broader definition that also includes respondents who report that they *could* get a loan if needed ("credit access").[10] It is important to note here that these questions were not targeted to input loans, and we have reason to believe that cash loans may work quite differently from input loans. Setting aside this concern for now, we interact these rudimentary measures of credit access at baseline with the treatment. Table 8 reports the results of this simple analysis on per-acre maize yields.

Columns (1) and (2) of Table 8 show results in the smaller, primary sample and columns (3) and (4) show the same regressions including the pre-exposed farmers. The latter two columns should be interpreted with some caution since the pre-exposed farmers' credit access may have been affected by their previous interactions with 1AF (if their previous participation was in 2016), making credit access and loan receipt bad controls in those regression. Focusing therefore on the first two columns, we can see that the treatment effects are very consistent with the main analysis in Section 4, but that farmers with baseline credit access seem to have lower yields for both measures of credit access. The interaction terms between baseline credit access and 1AF treatment status is positive but not significantly different from zero. This suggests that at least using this imperfect measure of credit access, treatment effects do not vary starkly with baseline credit access.

5.3.2 Indirect analysis of credit constraints: land assets

Given our concerns about the baseline credit variables, we look at the credit question from a few additional angles: first, we examine whether treatment effects vary by collateralizable assets, i.e., do treatment effects vary with baseline land access? Second, we examine currently-treated farmers' behaviors and yields on the land that they are farming but that they did not enroll in 1AF's program. If behaviors differ on treated and untreated plots, this may provide suggestive evidence weighing the relative importance of credit constraints versus information.

[10] The "reported loan use" question asked respondents "Have you received a loan or loans formally or informally other than from OAF in the past year?" Credit access is elicited by asking respondents who said they had *not* received a non-1AF loan in the past year, "...would [you] have been able to obtain a loan in the past year?"

Table 8: Yields per acre and credit

	Primary Sample (1)	(2)	Full Sample (3)	(4)
1AF participant	250.115***	318.829***	300.203***	305.722***
	(63.899)	(44.731)	(44.413)	(29.008)
Credit access 2016=1	-125.358**		-10.641	
	(48.377)		(31.000)	
1AF participant × Credit access 2016=1	84.943		-8.230	
	(64.819)		(46.372)	
Received loan at baseline=1		-105.039**		46.217
		(42.405)		(34.201)
1AF participant × Received loan at baseline=1		-21.874		-47.173
		(66.354)		(44.980)
Past 1AF participant=1			10.108	9.360
			(24.215)	(24.211)
Household size	10.891	9.077	7.072*	6.871*
	(7.617)	(7.501)	(4.108)	(4.117)
FAW Incidence	-5.533	-5.535	-3.681*	-3.669*
	(3.676)	(3.574)	(2.138)	(2.142)
Maize acres, 2016	44.946	48.417		
	(30.128)	(29.203)		
Maize yield/acre, 2016	0.142***	0.142***		
	(0.043)	(0.043)		
Observations	765	765	2137	2137
R^2	0.177	0.180	0.141	0.142
Ctrl Mean Dep. Var	1118.684	1118.684	1141.971	1141.971

Standard errors in parentheses

Note: standard errors clustered at farmer group level. The outcome variable is maize yields per acre. Results in this table are from OLS regressions which include field office area fixed effects.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 9: Yields per acre and baseline acres cultivated

	Primary Sample		Full Sample	
	(1)	(2)	(3)	(4)
1AF participant	313.485***	330.323***	356.884***	340.377***
	(67.815)	(54.271)	(38.472)	(32.847)
Maize acres, 2016	45.941		51.031**	
	(34.342)		(20.176)	
1AF participant × Maize acres, 2016	-3.722		-61.798**	
	(55.063)		(29.668)	
.75+ acres last season		54.310		102.347***
		(52.265)		(29.670)
1AF participant × .75+ acres last season		-43.573		-90.666**
		(69.647)		(43.180)
Past 1AF participant=1			9.852	9.180
			(24.378)	(24.312)
Married	129.328**	131.327**	57.856*	53.783*
	(58.850)	(59.120)	(31.940)	(31.860)
Father, 2ary school (0/1)	1.900	6.884	8.157	7.816
	(45.053)	(45.565)	(24.558)	(24.840)
Farm labor >50% income	36.674	40.645		
	(49.931)	(49.164)		
Used ag tech 2016	-38.391	-28.875		
	(39.529)	(40.465)		
Prev. 1AF knowledge	5.921	5.965		
	(65.357)	(65.013)		
Intercropped 2016	-23.909	-21.193		
	(43.629)	(42.993)		
Credit access 2016	-78.173**	-73.413**		
	(34.754)	(34.801)		
Observations	765	765	2137	2137
R^2	0.169	0.166	0.144	0.146
Ctrl Mean Dep. Var	1118.684	1118.684	1141.971	1141.971

Standard errors in parentheses

Note: standard errors clustered at farmer group level. The outcome variable is maize yields per acre. Results in this table are from OLS regressions which include field office area fixed effects.

* $p < .1$, ** $p < .05$, *** $p < .01$

In Table 9, we regress maize yields in 2017 on the treatment dummy, as well as an interaction with baseline acres planted in maize. Columns (1) and (3) show results interacting continuous maize acreage, while columns (2) and (4) use an indicator variable for planting more than three-quarters of an acre of maize. The continuous interactions are negative, but only statistically different from zero in the full sample where we might worry about past 1AF participation of some of the sample affecting how many acres farmers choose. Note, however, that the analysis in Section 4 may assuage concerns along these lines, as it did not reveal any impact of program participation on the number of acres planted in maize.

Similarly, the binary interactions have the signs that we would expect, but are only significant in the full sample. Figure 6 show the average marginal effect of treatment for farmers above and below the 0.75-acre threshold at baseline, and the associated 95% confidence intervals. These analyses provide some evidence that treatment has a larger impact on farmers who have smaller land endowments. While land assets at baseline may be correlated with credit access, we still want to caution the reader that this remains speculative.

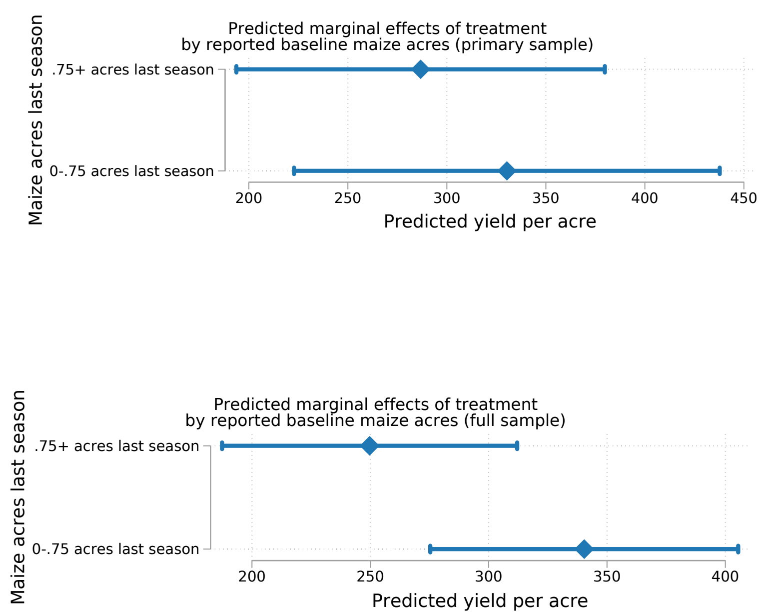


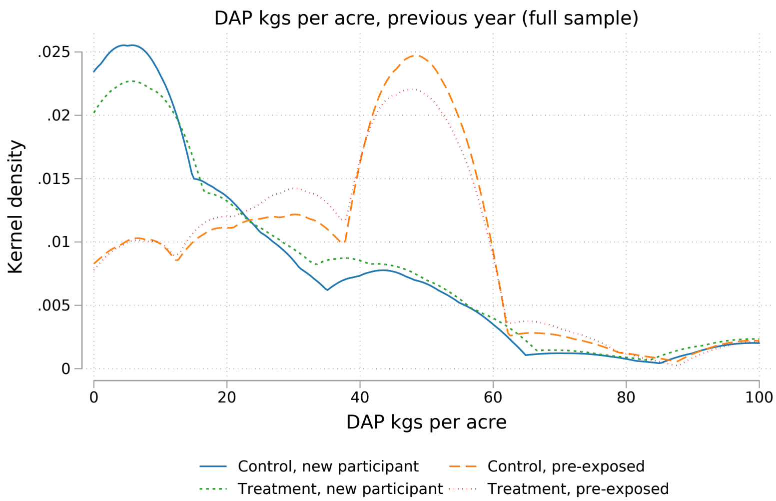
Figure 6: Marginal effect of treatment by baseline maize acres

5.3.3 Indirect analysis of credit constraints: fertilizer use

Another way that we can examine credit constraints is to return to input use. Section 5.2 evaluated whether or not input use defined as an either-or decision seemed relevant, but we have additional data on the usage intensity that we can exploit. Figures 1a and 1b shown in Section 2 showed that the unexposed control and treatment farmers look very similar in terms of the amount of fertilizer that they apply on their fields at baseline. If we additionally include pre-exposed farmers (Figure 7), we can observe that the pre-exposed are not only more likely to use fertilizer in general, they also use substantially more of both DAP and CAN than do control farmers. This figure is somewhat difficult to interpret, since many pre-exposed farmers were enrolled in the program during the baseline season, which could increase their fertilizer rates by the same mechanisms we see for new farmers in the study year. It could also be because these self-selected farmers were already using more fertilizer before ever enrolling.

Turning to the current season, we can also examine whether the fertilizer amounts used by pre-exposed and unexposed farmers varies once they are treated. Figure 8 splits the analysis by pre-enrollment status and divides a farmer's fields into enrolled and non-enrolled plots. We can see that the impact on fertilizer use is completely coming from the plots that are currently enrolled in the program. This holds for pre-exposed farmers and new farmers alike. Given that pre-exposed farmers presumably appreciated the program (after all, they chose to re-enroll!) why is there not more spillover onto their non-enrolled plots? It appears that the distributions for pre-exposed treated farmers' non-enrolled plots lie slightly to the right of the new treated farmers' intensity distribution, but we cannot reject that these are drawn from the same distribution.

Perhaps the most compelling explanation for the lack of spillovers between enrolled and non-enrolled plots is that of credit constraints. If farmers can only afford to purchase the fertilizer amounts that the program provides them with on loan, then they may be unable to afford to bring their other fields up to the recommended fertilizer intensity levels. Granted, alternative explanations abound: farmers may choose to enroll only the plot that they believe will have the highest returns to the program. We have little ability to examine this hypothesis directly. Yet another explanation could be that the inputs that 1AF provides are of higher quality, and farmers do not want to spend money on inferior inputs. As Bold et al. (2017); Tjernström et al. (2018) have demonstrated,



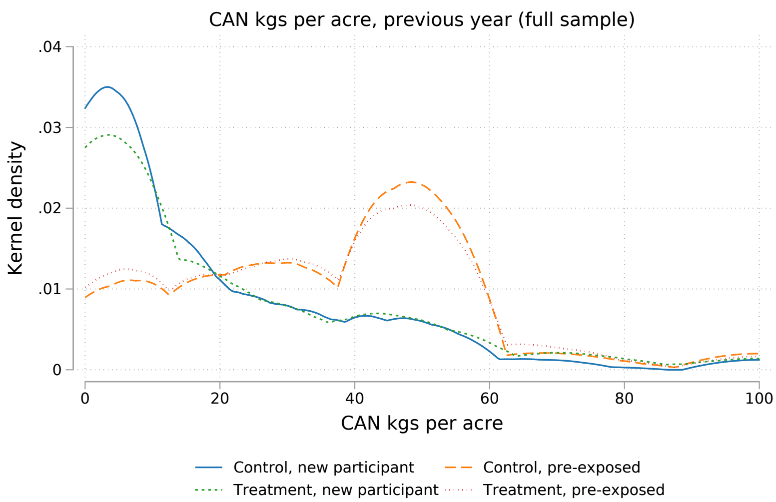
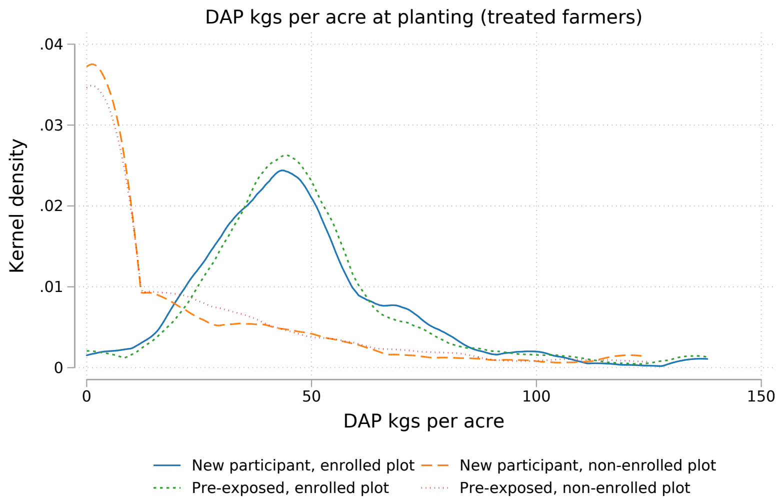


Figure 7: Fertilizer intensity, baseline



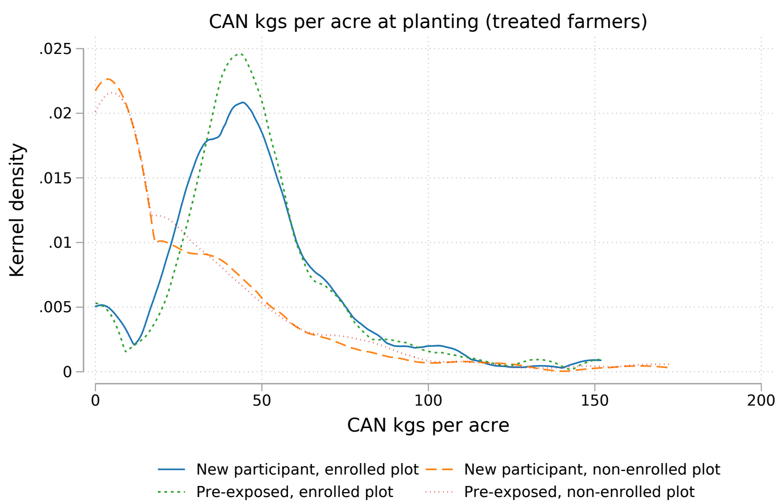


Figure 8: Fertilizer intensity, current season

fertilizer is often missing advertised nutrient levels.

We can, however, examine whether there appears to be a cumulative effect of repeated enrollment in the program. If the intervention acts as a graduation program, building farmers up to a higher level of productivity, we might expect to see changes in the intensive margin of program participation over time. The next section studies this dimension, using past enrollment and future enrollment information.

5.4 Re-enrollment and effects of cumulative enrollment

We now try to understand how farmers' past participation relates to current outcomes, and the factors that correlate with farmers' extensive and intensive enrollment decisions. To answer this question, we match farmers to 1AF administrative program data which includes information about farmers' enrollment history and loan repayment.

Of particular interest is how farmers choose to re-enroll in the program. We do not have strong *a priori* expectations about the extensive margin effect. If farmers value the program because of the resources that it provides, farmers may be more likely to re-enroll if they participated in the previous season. However, if the program is truly effective at helping farmers "graduate" to higher yields, then we might expect some treated farmers to drop out of the program if they no longer feel they need it. In Table 10 we see that being randomly allocated to participation in 2017 did not significantly increase the probability that farmers enrolled in 2018. However, we do see that past participation is a positive and significant predictor of 2018 enrollment. This suggests there may have been some self-selection into earlier participation by farmers who appreciate the program more. This should, if anything, strengthen our confidence in the main program effects for the primary sample, since the randomization occurred over the non-self-selected farmers.

In order to think about potential "graduation" effects, we turn to the question of how much land farmers choose to enroll. If farmers find it useful to enroll more land to access more credit and/or larger quantities of high-quality inputs, then they may increase enrollment year-on-year after their first year of participation. However, if the program is more useful to farmers for its information effects, then we may not expect any change in land enrolled. We know from Table 1 that farmers in treatment and control groups were not significantly different in baseline maize acres cultivated. We might therefore expect that treated and control farmers on average would choose to enroll the same

Table 10: Enrollment in 2018, primary and full samples

	(1) Enrolled 2018 (primary sample)	(2) Enrolled 2018 (full sample)
1AF participant	0.000	-0.004
	(0.030)	(0.019)
Past 1AF participant		0.056***
		(0.017)
Married	0.029	0.005
	(0.047)	(0.024)
Father, 2ary school (0/1)	-0.005	-0.021
	(0.035)	(0.015)
Farm labor >50% income	0.037	
	(0.037)	
Used ag tech 2016	0.028	
	(0.029)	
Prev. 1AF knowledge	0.091*	
	(0.046)	
Intercropped 2016	-0.009	
	(0.029)	
Credit access 2016	-0.035	
	(0.034)	
Household size	0.011**	0.004
	(0.005)	(0.003)
FAW Incidence	0.001	-0.001
	(0.003)	(0.002)
Maize acres, 2016	-0.006	
	(0.022)	
Maize yield/acre, 2016	-0.000	
	(0.000)	
Observations	765	2137
R^2	0.350	0.442
Control Mean Dep. Var	0.662	0.672

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions which include field office area fixed effects. The outcome variable is a binary indicator of whether the farmer appeared in the 2018 administrative data as enrolled in the program.

* $p < .1$, ** $p < .05$, *** $p < .01$

amount of land in the program in 2017.

Table 11 shows that being randomly assigned to participate in the program in 2017 significantly increased land enrolled in 2018. Column (1) includes only the sample of farmers who did choose to re-enroll, whereas column (2) adds farmers who did not re-enroll by treating their enrolled acres as zero. This second column should be a very conservative estimate of the effect. This lends further support to the idea that farmers find the access to credit and quality inputs useful. We might also interpret these results as suggesting that program participation in one year may be easing constraints, allowing farmers to increase enrollment in the next year. Alternatively, it could suggest a learning mechanism, whereby farmers are more interested in enrolling more of their land having experimented with the program in the past year.

We can go further to attempt to distinguish between eased constraints and farmer learning as explanations for increased land enrollment by treatment farmers. In column (1) of Table 12, we show enrollment decisions in 2018 by farmers who joined 1AF in 2016. Although initial enrollment was non-random, control farmers were randomly held out from the program in 2017. Thus, exposure to one vs two years of the program is random. The effect of 2017 participation on 2018 acres enrolled is positive and significant. This sample was already aware of how the program worked, having participated in 2016, so we take this as evidence that the mechanism that leads farmers to increase enrollment in the next year runs through a relaxation of constraints (by increasing yields and profits). We do see a positive point estimate for randomized exposure to a third year of participation for 2015 enrollees, but the effect is not precisely estimated. This could be due to sample size issues, but could alternatively suggest that multi-year participants face other constraints that prevent them from increasing land enrollment, like cultivable land or labor availability.

5.5 Heterogeneous Treatment Effects

In the subsections above we considered several key dimensions of interest along which we might expect the effect of 1AF participation to vary. However, we are also interested in understanding more generally whether there was substantial heterogeneity in the effect of program participation, and whether key observable variables can tell us anything about farmers who benefited the most or the least.

To answer this question, we apply new machine learning (ML) methods from Chernozhukov

Table 11: Acres enrolled in 2018, primary sample

	(1) 2018 Acres	(2) 2018 Acres, Missings=0
1AF participant	0.130***	0.077**
	(0.038)	(0.035)
Married	-0.000	0.019
	(0.054)	(0.044)
Father, 2ary school (0/1)	0.009	-0.004
	(0.030)	(0.030)
Farm labor >50% income	-0.050	-0.008
	(0.043)	(0.035)
Used ag tech 2016	0.013	0.039
	(0.029)	(0.031)
Prev. 1AF knowledge	0.109*	0.142***
	(0.064)	(0.050)
Intercropped 2016	-0.045	-0.031
	(0.032)	(0.027)
Credit access 2016	0.067**	0.023
	(0.029)	(0.028)
Household size	0.004	0.009*
	(0.006)	(0.006)
FAW Incidence	-0.000	0.000
	(0.003)	(0.003)
Maize acres, 2016	0.106***	0.064**
	(0.022)	(0.027)
Maize yield/acre, 2016	0.000**	0.000
	(0.000)	(0.000)
Observations	490	765
R^2	0.222	0.266
Control Mean Dep. Var	0.617	0.409

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions which include field office area fixed effects. Column (2) fills missing 2018 enrolled acres as zero, so includes farmers who chose not to re-enroll.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 12: Acres enrolled in 2018, from past new enrollees

	(1) 2018 Acres (2016 Enrollees)	(2) 2018 Acres (2015 Enrollees)
1AF participant	0.220***	0.064
	(0.040)	(0.042)
Married	0.045	0.009
	(0.050)	(0.061)
Father, 2ary school (0/1)	0.071*	0.053
	(0.038)	(0.040)
Household size	0.018***	0.026**
	(0.006)	(0.010)
FAW Incidence	-0.001	-0.003
	(0.004)	(0.004)
Observations	397	293
R^2	0.200	0.170
Control Mean Dep. Var	0.580	0.625

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions which include field office area fixed effects.

* $p < .1$, ** $p < .05$, *** $p < .01$

et al. (2018) to conduct inference on a few key features of heterogeneous treatment effects.[11] 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.

Briefly, the method throughout 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

[11] Our discussion of the method draws heavily on the discussion in Chernozhukov et al. (2018). Section 6.2 in their paper describes the implementation algorithm in detail.

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 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:

$$Y_i = \hat{\alpha}' X_{1,i} + \hat{\beta}_1(D_i - p(Z_i)) + \hat{\beta}_2(D_i - p(Z_i))(S_i - \mathbb{E}_N S_i) + \hat{\varepsilon}_i \tag{3}$$

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 $S(Z_i)$. The purpose of writing this out is to hopefully provide a bit more intuition for the results in Table 13. 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 $\beta_2 = 0$. This does not tell us with certainty that there exists no heterogeneity in treatment effects, but it does tell us that the vector of covariates included has no power to predict treatment effect heterogeneity.

The results presented in this section all use the primary sample. We include as explanatory variables the full set of controls from Table 1, as well as several additional baseline control variables: distance to nearest farmer group member, acres planted in the short rains season in 2016, per-acre

CAN and DAP intensity, per-acre use of commercial improved seeds, number of extension officer visits, and indicator variables for whether the household hired any labor, planted diverse crops, claimed knowledge of 1AF practices, had a homestead (not household) member who had ever enrolled in 1AF, was a member of other farming organizations, had electricity access, and saved any money in the year before the baseline.

We first report the estimated Conditional Average Treatment Effect (CATE, corresponding to β_1 in Eq. 3) and Heterogeneity Loading Parameters (HET, corresponding to β_2 in Eq. 3)) from two ML methods estimators in Table 13. The listed ML methods are those that were selected based on the methods' relative performance across the various splits of the data. In our case, the program selects neural net and random forest methods as the best-performing. In columns (1) and (3), we find similar magnitudes for our main treatment effects as with standard regression analysis, although we lose some precision with the ML approach. In particular, we fail to reject that the profit effect is different from zero. This is perhaps expected, since the sampling noise with these methods can be substantial when there is little heterogeneity (that said, the post-processing of the ML estimates should in theory help reduce the noise).

Columns (2) and (4) of Table 13 show that we are unable to reject the null hypothesis that the heterogeneity loading parameter is zero, i.e. the covariates that we include here are not predictive of treatment effect heterogeneity. The confidence intervals on the HET estimates are not very tight, but they appear to be centered around zero and are certainly far from 1. This suggests that our estimated proxy for heterogeneity explains very little of any actual heterogeneity in the treatment effects. This could be because we are using the wrong covariates, and in future work we aim to include a broader set of covariates, as well as interactions between covariates and weather realizations.

To construct groups for predicting GATEs, we bin observations by predicted actual treatment effect into five groups. Figures 9, 10, and 11 show the resulting predicted average treatment effect by group with maize per acre, maize per farmer, and profit as the outcomes, respectively. In each figure, we display the two methods that performed best according to the guidelines in Chernozhukov et al. (2018). In our case, this turns out to be neural nets (in the left hand side panels) and random forests (in the right hand side panels). These figures overall corroborate the notion that treatment effect heterogeneity due to the included covariates is rather limited, since the GATES point estimates of

Table 13: BLP of Conditional Average Treatment Effect by outcome, primary sample

	Neural Net		Random Forest	
	CATE	HET	CATE	HET
Profit	51.13	0.035	46.22	0.026
	(-32.52,139.9)	(-0.298,0.372)	(-39.87,128.5)	(-0.262,0.323)
	[0.511]	[1.000]	[0.618]	[1.000]
Yields Per Acre	308.2	0.013	307.9	-0.058
	(196.8,417.3)	(-0.295,0.352)	(193.6,419.6)	(-0.335,0.214)
	[0.000]	[1.000]	[0.000]	[1.000]
Yields Per Farmer	257.2	-0.003	229.7	-0.002
	(24.28,498.2)	(-0.382,0.363)	(4.900,469.5)	(-0.326,0.315)
	[0.063]	[1.000]	[0.093]	[1.000]

Note: Medians over 100 sample splits. 90% confidence interval reported in parentheses.

the highest and lowest groups are barely outside the original confidence intervals of the original ATE point estimate (the dashed red lines).

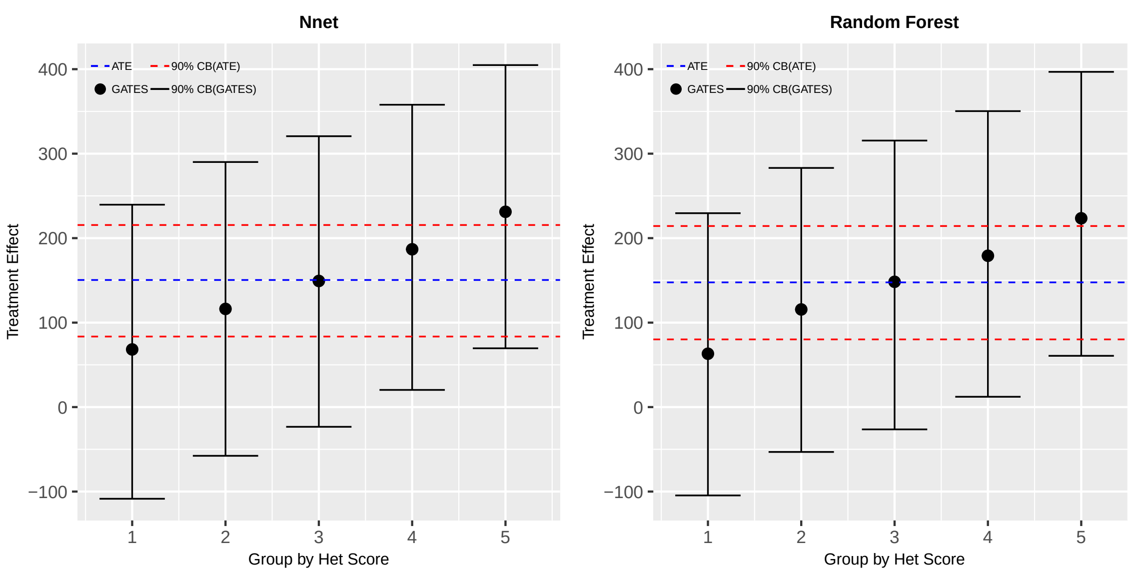


Figure 9: Group Average Treatment Effects for Maize Yields per Acre, primary sample

Next we turn to results of the Classification Analysis, seen in Table 14. Here we focus on the left-most and right-most groups in Figure 9. Note that the two methods differ somewhat in terms of the estimated heterogeneity. While the Neural Net (columns (1)-(3)) fails to detect any significant differences in means between the most- and least-affected groups, the Random Forest method (columns (4)-(6)) finds some weak evidence. In particular, it suggests that the most-affected group have on average larger households than the least-affected group, and the two groups appear to

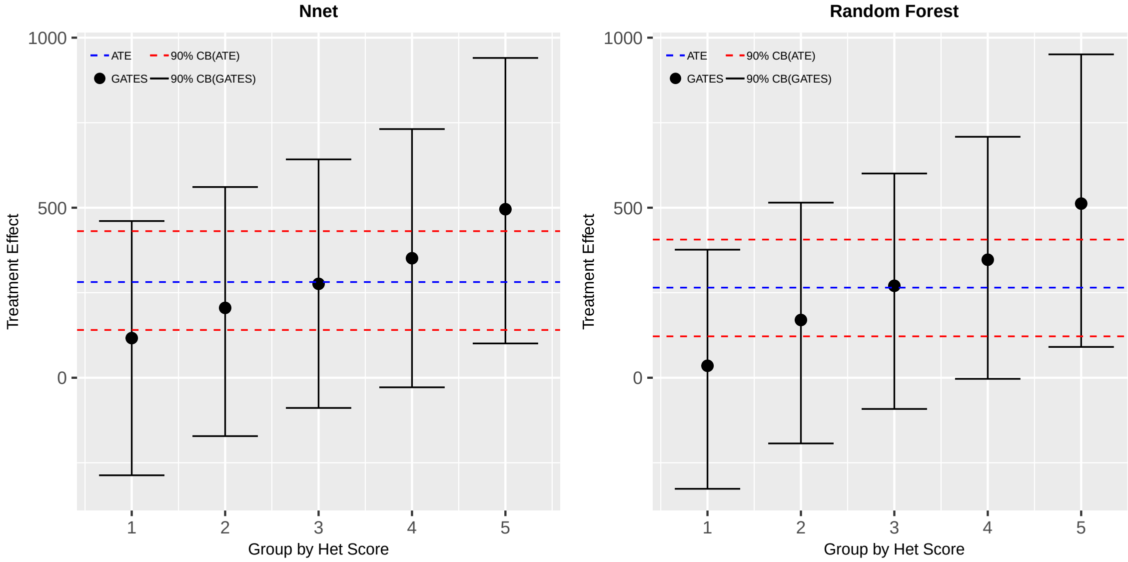


Figure 10: Group Average Treatment Effects for Maize Yields per Farmer, primary sample

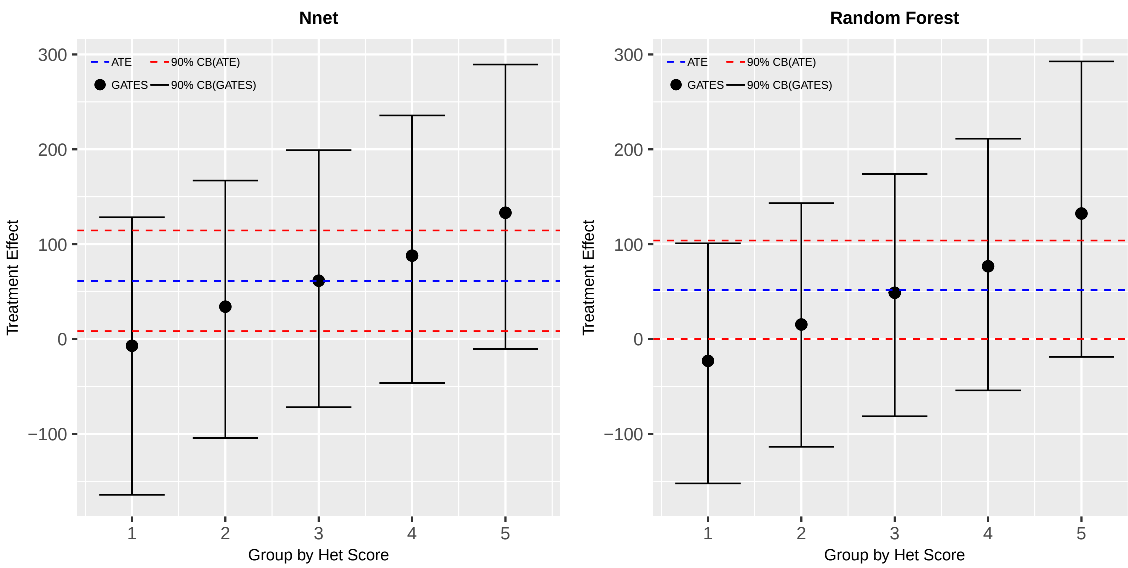


Figure 11: Group Average Treatment Effects for Profit, primary sample

differ on other dimensions related to baseline practices as well. The most-affected group (i.e., those with the highest predicted treatment effects) applied less fertilizer and improved seeds per acre at baseline. We do not want to read too much into this result, since we know from above that overall our covariates seem to have little predictive power for treatment effect heterogeneity. If we do want to interpret the results in columns (4)-(6), it is a rather intuitive result: farmers who were already using improved seeds and more fertilizer may not have benefited as much from the program as they already had better information and/or resources.

Table 14: Classification Analysis comparing baseline characteristics of most- and least-affected groups (outcome variable: yields per acre; sample: primary)

	Neural Net			Random Forest		
	Most Affected	Least Affected	Difference	Most Affected	Least Affected	Difference
Household Size	6.678	6.474	0.281	7.153	6.404	0.786
	(5.964,7.458)	(5.783,7.296)	(-1.028,1.283)	(6.486,7.843)	(5.762,7.090)	(-0.095,1.702)
	-	-	[1.000]	-	-	[0.156]
Credit Access	0.692	0.687	0.021	0.708	0.687	0.034
	(0.569,0.834)	(0.563,0.809)	(-0.187,0.202)	(0.589,0.832)	(0.571,0.808)	(-0.140,0.195)
	-	-	[1.000]	-	-	[1.000]
Maize acres, 2016	1.068	0.973	0.098	1.188	0.936	0.225
	(0.850,1.287)	(0.754,1.182)	(-0.211,0.415)	(0.983,1.391)	(0.728,1.137)	(-0.049,0.505)
	-	-	[0.887]	-	-	[0.231]
CAN per acre, 2016	14.470	19.78	-5.416	12.320	22.88	-11.180
	(8.578,21.10)	(13.91,25.96)	(-14.15,3.220)	(6.721,18.03)	(17.07,28.22)	(-19.08,-3.253)
	-	-	[0.409]	-	-	[0.014]
DAP per acre, 2016	18.620	25.45	-6.222	16.020	27.66	-11.490
	(10.22,25.95)	(17.49,32.72)	(-16.69,4.718)	(9.301,22.41)	(21.37,34.22)	(-20.74,-2.115)
	-	-	[0.541]	-	-	[0.034]
Commercial seeds per acre, 2016	5.278	6.762	-1.490	4.932	6.918	-2.013
	(3.922,6.452)	(5.582,7.898)	(-3.436,0.265)	(3.845,6.042)	(5.789,7.989)	(-3.525,-0.487)
	-	-	[0.201]	-	-	[0.019]

Note: Medians over 100 sample splits. 90% confidence interval reported in parentheses. Most affected and least affected are groups 1 and 5 on the above figures.

6 Robustness checks

In this section, we explore the robustness of our primary results to different samples, specifications, and types of inference. Table 15 summarizes the various robustness checks and their results. We present some of the tables with results in this section, and others can be found in the relevant appendices.

Table 16 estimates the effect of 1AF participation on our three main outcomes in the full sample, i.e., including pre-exposed farmers (compare with our preferred specification in Table 3). In columns (1) and (2), we can see that the per-acre and per-farmer maize yield impacts are very similar to

Table 15: Robustness Checks

Robustness Check	Result
Primary regressions with full sample	Decreased magnitude, no change to significance (see Table 16)
Drop controls and drop fixed effects in primary and full samples	Changes in magnitude, no change to significance (see Tables C.1 and C.2)
Correct for multiple hypothesis testing using Westfall and Young (1993)	No changes to significance
Randomization inference on all outcomes with primary and full samples	See Figures 12 and 13
Impute missing non-program yields	Decreased magnitudes, decreased significance of treatment effect estimate on profit to 10% level (see Tables C.3, C.5, C.4, and C.6)

those using the primary sample. The results in column (3) show that treatment impacts on farm profits are a bit smaller than those using the primary sample, but not by very much. Furthermore, all of the results remain statistically different from zero and economically meaningful. This suggests that including the pre-exposed farmers does not dramatically bias the estimated treatment effects.

To assure the reader that the inclusion of covariates in our preferred specifications is not biasing our results, Appendix Table C.1 presents the primary analysis (compare with Table 3) including only the treatment dummy (columns 1, 3, and 5) as well as regressions excluding covariates but including site fixed effects (columns 2, 4, and 6). The stability of the treatment effects across the various specifications should help persuade skeptical readers that covariate adjustment is not driving our results.

To prod the fidelity of our inference, we take a leaf from Young (2017) and implement a randomization inference procedure using Heß (2017). For each regression, we randomly permute the treatment dummy (maintaining cluster-level treatment assignment) 5,000 times, calculating the difference in means between treatment and control for each permutation. This creates an approximation of the reference distribution under the sharp null hypothesis of no treatment for any participant.[12] Just as adjusting for multiple hypothesis testing did not change the statistical significance of our results (adjusted p-values reported at the bottom of Table 3), our results remain statistically significant under randomization inference. Figures 12-13 show the density of

[12] It is an approximation because we are not computing the complete set of possible permutations.

Table 16: Primary outcomes, full sample

	(1) Maize PA	(2) Maize Per Farmer	(3) Profit
1AF participant	289.047***	281.794***	52.637***
	(23.899)	(49.697)	(18.391)
Past 1AF participant	-24.426	3.233	3.330
	(26.341)	(63.445)	(22.680)
Married	62.497*	40.816	3.163
	(31.626)	(73.664)	(26.931)
Father, 2ary school (0/1)	0.446	203.929***	68.691***
	(24.526)	(51.499)	(18.979)
Farm labor >50% income	46.799*	102.909*	43.303**
	(26.581)	(55.483)	(20.874)
Used ag tech 2016	-28.172	-44.779	-24.977
	(33.354)	(67.917)	(24.957)
Prev. 1AF knowledge	21.266	63.991	12.627
	(28.635)	(55.526)	(21.047)
Intercropped 2016	-12.863	-52.260	-10.987
	(24.763)	(58.359)	(21.446)
Credit access 2016	-22.207	-27.312	-20.923
	(22.270)	(52.687)	(19.244)
Observations	2137	1785	1785
R^2	0.157	0.225	0.178
MHT p-value for 1AF	0.000	0.000	0.005
Control Mean Dep. Var	1141.971	1154.645	364.480

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions which include field office area fixed effects. Columns (2) and (3) exclude observations missing non-program yields (see Appendix C). MHT p-value uses Jones et al. (2018) implementation of Westfall and Young (1993) correction for multiple hypothesis testing.

* $p < .1$, ** $p < .05$, *** $p < .01$

post-permutation treatment effect estimates (solid lines) compared to the estimate from the true assignments (the dashed lines). The densities is our approximation of the "null distribution", i.e. the distribution of treatment effects under the sharp null of no treatment effect for anyone. By comparing our treatment effect estimates (dashed lines) to these densities, it becomes clear that it is quite unlikely that our treatment effects would have occurred by chance in a random permutation of the treatment assignment vector.

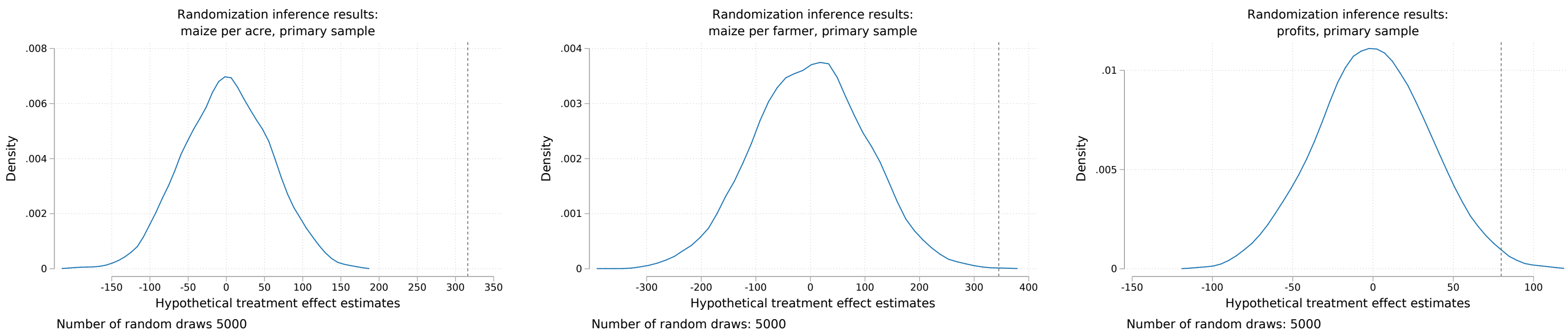


Figure 12: Randomization inference results with primary sample

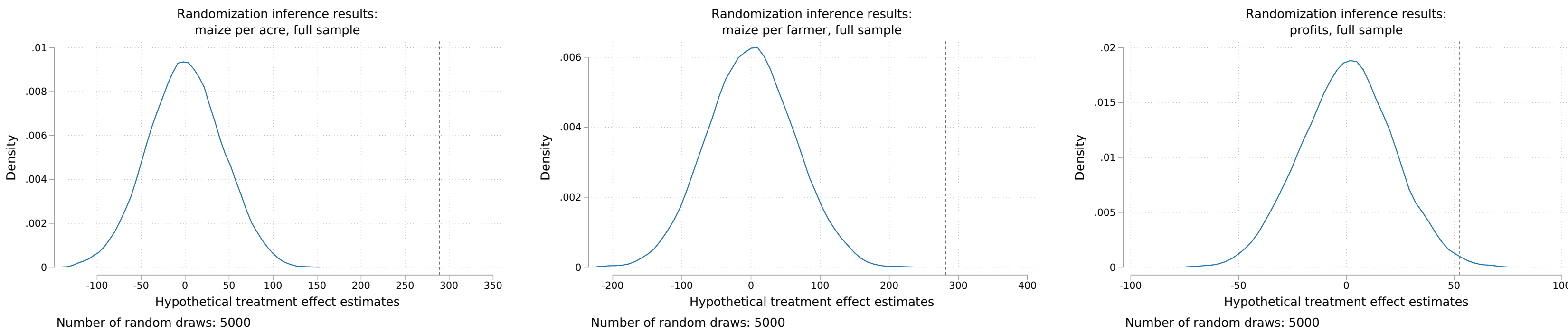


Figure 13: Randomization inference results with full sample

7 Discussion and conclusions

This paper presents evidence that a bundled intervention targeted at smallholder farmers effectively increases smallholder productivity in a region where few interventions have succeeded at this goal. We demonstrate that the effect is robust to a variety of specifications, robustness checks, and inference that adjusts for multiple hypothesis testing. Given the number of interventions that have unsuccessfully attempted to improve farmer yields and incomes in sub-Saharan Africa, the fact that the One Acre Fund has achieved scaled-up success may appear as a puzzle to some readers. Our results suggest that the relaxation of multiple constraints at a time, consistent with poverty traps theory, may play an important role. Furthermore, the relatively low levels of treatment impact heterogeneity suggest that the program rarely fails.

To dig into the constraints that farmers might face, we use various approaches to heterogeneity analysis and find evidence that the target population face multiple constraints such that relaxing any single constraint (such as information) unlikely explains the 1AF program's success. We uncover suggestive evidence that more disadvantaged farmers (those who have less intensive agricultural practices at baseline) benefit more from participation. We also find suggestive evidence of credit constraints, albeit somewhat speculative as we do not have convincing direct measures of credit status. Our machine learning results are unable to reveal substantial heterogeneity, at least given the set of covariates that we observe. The limited heterogeneity that we do find in the Classification Analysis in Section 5.4 corroborates the notion that those farmers with greater constraints in the baseline experience the largest improvements. We additionally find that farmers increase their participation over the intensive margin over time. These results are in line with the literature on graduation programs, but instead within the context of agriculture rather than entrepreneurship programs (the typical focus of anti-poverty graduation programs).

While we do not directly observe income dynamics, and while our study duration is relatively short, the evidence that we uncover is consistent with the large literature on poverty traps. In the models common to that literature, multiple constraints act simultaneously to keep the poor in poverty. In poverty trap contexts, it is only possible to substantially improve long-term outcomes by relaxing multiple constraints at once, allowing households to get on to a higher-potential trajectory.

To fully understand the relative impacts of the various components of the 1AF program, we would have had to run a different study that provided subsets of the program to some farmers. This would also have allowed for a more direct cost-effectiveness analysis: could a leaner version of the 1AF core program achieve nearly the same impacts as this bundled version? However, the fact that multiple markets appear to be functioning imperfectly in this context provides additional motivation for offering a bundled program that solves issues in multiple markets at once.

References

Ashraf, N., X. Giné, and D. Karlan (2009): "Finding Missing Markets (and a Disturbing Epilogue): Evidence from an Export Crop Adoption and Marketing Intervention in Kenya," *American Journal of Agricultural Economics*, 91, 973–990.

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

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Pariente, J. Shapiro, B. Thuysbaert, and C. Udry (2015): "A multifaceted program causes lasting progress for the very poor: Evidence from six countries," *Science*, 348, 1260799–1260799.

Beegle, K., L. Christiaensen, A. Dabalen, and I. Gaddis (2016): *Poverty in a Rising Africa*, The World Bank.

Benjamin, D. (1992): "Household composition, labor markets, and labor demand: testing for separation in agricultural household models," *Econometrica: Journal of the Econometric Society*, 287–322.

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

Carter, M. R., R. Laajaj, and D. Yang (2013): "The Impact of Voucher Coupons on the Uptake of Fertilizer and Improved Seeds: Evidence from a Randomized Trial in Mozambique," *American Journal of Agricultural Economics*, 95, 1345–1351.

——— (2014): "Subsidies and the persistence of technology adoption: Field experimental evidence from Mozambique," Working Paper, NBER.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018): "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments," Working Paper, NBER.

Conley, T. G. and C. R. Udry (2010): "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, 100, 35–69.

Deutschmann, J. W. and E. Tjernström (2018): "The impact of One Acre Fund's small farm program," Technical Report.

Dillon, B. and C. B. Barrett (2017): "Agricultural factor markets in Sub-Saharan Africa: An updated view with formal tests for market failure," *Food Policy*, 67, 64.

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

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

Elabed, G. and M. Carter (2014): "Ex-ante impacts of agricultural insurance: Evidence from a field experiment in Mali," Working paper.

Feder, G., R. E. Just, and D. Zilberman (1985): "Adoption of Agricultural Innovations in Developing Countries: A Survey," *Economic Development and Cultural Change*, 33, 255–298, articleType: research-article / Full publication date: Jan., 1985 / Copyright © 1985 The University of Chicago Press.

Giné, X. and D. Yang (2009): "Insurance, credit, and technology adoption: Field experimental evidence from Malawi," *Journal of Development Economics*, 89, 1–11.

Gollin, D., C. W. Hansen, and A. Wingender (2018): "Two blades of grass: The impact of the green revolution," Working Paper, NBER.

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

Hanna, R., S. Mullainathan, and J. Schwartzstein (2014): "Learning Through Noticing:

Theory and Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 129, 1311–1353.

Hess, S. H. (2017): "Randomization inference with Stata: A guide and software," *Stata Journal*, 17, 630–651.

Intermedia Development Consultants (2017): "Review of One Acre Fund Data Collection: Step 3 Report," Field Report.

Jones, D., D. Molitor, and J. Reif (2018): "What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study," Working Paper, NBER.

Karlan, D., E. Kutsoati, M. McMillan, and C. Udry (2011): "Crop Price Indemnified Loans for Farmers: A Pilot Experiment in Rural Ghana," *Journal of Risk and Insurance*, 78, 37–55.

Karlan, D., R. Osei, I. Osei-Akoto, and C. Udry (2014): "Agricultural Decisions after Relaxing Credit and Risk Constraints," *The Quarterly Journal of Economics*, 129, 597–652.

Kirimi, L., N. Sitko, T. S. Jayne, F. Karin, M. Muyanga, M. Sheahan, J. Flock, and G. Bor (2011): "A farm gate-to-consumer value chain analysis of Kenya's maize marketing System," Technical Report WPS 44/2011, Tegemeo Institute of Agricultural Policy and Development.

LaFave, D. and D. Thomas (2016): "Farms, families, and markets: New evidence on completeness of markets in agricultural settings," *Econometrica*, 84, 1917–1960.

Ligon, E. and E. Sadoulet (2008): "Estimating the effects of aggregate agricultural growth on the distribution of expenditures," Background Paper, World Bank, Washington, D.C.

Maertens, A., H. Michelson, and V. Nourani (2017): "How do Farmers Learn from Extension Services? Evidence from Malawi," Working Paper.

Magruder, J. R. (2018): "An Assessment of Experimental Evidence on Agricultural Technology Adoption in Developing Countries," *Annual Review of Resource Economics*.

Mobarak, A. and M. Rosenzweig (2012): “Selling formal insurance to the informally insured,” Working Paper.

Sheahan, M., R. Black, and T. S. Jayne (2013): “Are Kenyan farmers under-utilizing fertilizer? Implications for input intensification strategies and research,” *Food Policy*, 41, 39–52.

Tjernström, E., M. R. Carter, and T. Lybbert (2018): “The dirt on dirt: Soil characteristics and variable fertilizer returns in Kenyan maize systems,” Working Paper.

Udry, C. (1996): “Efficiency and Market Structure: Testing for Profit Maximization in African Agriculture,” Tech. rep.

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

Young, A. (2017): “Channelling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results,” Working Paper.

Appendix

A Pre-Registered Analyses

In this section we review the main commitments in the pre-analysis plan, and note any deviations. In terms of data composition, our samples are slightly smaller than expected. The primary culprit is missing data on non-program-land yields by enrolled farmers. This issue is addressed more fully below in Appendix C.

The pre-analysis plan specifies three outcomes of interest: maize yields per acre, maize yields per farmer, and profit. The planned specification includes treatment status, and the following controls: marital status, household size, landownership size, education, agricultural reliance, credit constraints, use of agricultural tech in the last season, intercropping, and knowledge of 1AF practices. Where appropriate, regressions should also include a control for whether the farmer was pre-exposed. Additionally, the PAP planned to include a spillover inverse probability weight. Our results are robust to including this weighting, but for simplicity we present all results without weights. The PAP specified additional analysis to explore "enduring impacts" using only the control sample, i.e. exploring whether the pre-exposed farmers in the control group were still better off.

Table	Relation to PAP
3	Controls match PAP
4	Not mentioned in PAP
5	Not mentioned in PAP
6	Specified in PAP but includes more limited controls
7	Not mentioned in PAP
8	Not mentioned in PAP
9	Not mentioned in PAP
16	Controls match PAP
C.1	Outcomes match PAP but shown without controls

Table A.1: Tables and PAP

B Balance tables for primary and pre-exposed separately

	(1) Control		(2) Treatment		Difference
Variable	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	(1)-(2)
Married	399	0.902	366	0.883	0.020
	[58]	(0.016)	[59]	(0.016)	
Father, 2ary school (0/1)	399	0.348	366	0.377	-0.029
	[58]	(0.031)	[59]	(0.026)	
Farm labor >50% income	399	0.789	366	0.751	0.038
	[58]	(0.032)	[59]	(0.026)	
Used ag tech 2016	399	0.619	366	0.653	-0.034
	[58]	(0.037)	[59]	(0.031)	
Prev. 1AF knowledge	399	0.063	366	0.142	-0.079***
	[58]	(0.015)	[59]	(0.021)	
Intercropped 2016	399	0.476	366	0.557	-0.081
	[58]	(0.034)	[59]	(0.040)	
Credit access 2016	399	0.707	366	0.708	-0.001
	[58]	(0.029)	[59]	(0.028)	
Household size	399	6.624	366	6.661	-0.037
	[58]	(0.151)	[59]	(0.155)	
Maize acres, 2016	399	1.003	366	0.945	0.058
	[58]	(0.064)	[59]	(0.054)	
Maize yield/acre, 2016	399	429.007	366	445.686	-16.679
	[58]	(32.729)	[59]	(36.551)	
F-test of joint significance (F-stat)					1.834*
Number of observations					765

Notes: Field office fixed effects are included in all estimation regressions. Standard errors clustered at farmer group cluster level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table B.1: Baseline balance (primary sample)

	(1)		(2)		
	Control		Treatment		Difference
Variable	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	(1)-(2)
Married	663	0.863	709	0.879	-0.016
	[60]	(0.015)	[59]	(0.014)	
Father, 2ary school (0/1)	663	0.410	709	0.453	-0.042
	[60]	(0.024)	[59]	(0.030)	
Farm labor >50% income	663	0.772	709	0.793	-0.020
	[60]	(0.021)	[59]	(0.020)	
Used ag tech 2016	663	0.881	709	0.879	0.002
	[60]	(0.014)	[59]	(0.014)	
Prev. 1AF knowledge	663	0.710	709	0.722	-0.012
	[60]	(0.024)	[59]	(0.024)	
Intercropped 2016	663	0.478	709	0.433	0.045
	[60]	(0.032)	[59]	(0.031)	
Credit access 2016	663	0.713	709	0.733	-0.020
	[60]	(0.023)	[59]	(0.023)	
Household size	663	6.674	709	6.906	-0.231
	[60]	(0.122)	[59]	(0.112)	
Maize acres, 2016	663	1.008	709	1.066	-0.058
	[60]	(0.039)	[59]	(0.040)	
Maize yield/acre, 2016	663	597.347	709	650.686	-53.339**
	[60]	(32.232)	[59]	(30.494)	
F-test of joint significance (F-stat)					1.375
Number of observations					1372

Notes: Field office fixed effects are included in all estimation regressions. Standard errors clustered at farmer group cluster level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table B.2: Baseline balance (pre-exposed sample)

C Sample definition

We define the sample used in our primary regressions above to be farmers for which we have a non-missing primary outcome (maize yields per acre) and non-missing values for the pre-specified controls. Here we attempt to address several potential concerns with this sample definition. First, in Tables C.1 and C.2, we demonstrate that our coefficients of interest are similar when we exclude controls and use the full sample with non-missing outcomes, whether or not we also include field office fixed effects.

Table C.1: Primary outcomes without controls, primary sample

	Maize PA		Maize Per Farmer		Profit	
	(1)	(2)	(3)	(4)	(5)	(6)
1AF participant	302.875***	293.733***	376.650***	335.889***	92.273**	76.321**
	(47.437)	(34.904)	(131.786)	(95.411)	(45.663)	(33.358)
Observations	828	828	679	679	679	679
R^2	0.081	0.151	0.028	0.148	0.013	0.129
MHT p-value for 1AF	0.000	0.000	0.005	0.001	0.046	0.024
Control Mean Dep. Var	1124.733	1124.733	1081.431	1081.431	337.700	337.700
Site FE	N	Y	N	Y	N	Y

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions. Columns 3-6 exclude observations missing non-program yields (see Appendix C). MHT p-value uses Jones et al. (2018) implementation of Westfall and Young (1993) correction for multiple hypothesis testing.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table C.2: Primary outcomes without controls, full sample

	Maize PA		Maize Per Farmer		Profit	
	(1)	(2)	(3)	(4)	(5)	(6)
1AF participant	290.345***	290.308***	391.616***	348.123***	88.053***	73.587***
	(33.181)	(23.215)	(79.999)	(52.418)	(27.671)	(18.839)
Past 1AF participant	19.896	9.519	113.887*	146.436**	35.500	42.611**
	(24.387)	(23.489)	(64.747)	(60.782)	(23.021)	(21.472)
Observations	2251	2251	1859	1859	1859	1859
R^2	0.079	0.135	0.033	0.100	0.014	0.078
MHT p-value for 1AF	0.000	0.000	0.000	0.000	0.002	0.000
Control Mean Dep. Var	1143.077	1143.077	1145.997	1145.997	361.164	361.164
Site FE	N	Y	N	Y	N	Y

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions. Columns 3-6 exclude observations missing non-program yields (see Appendix C). MHT p-value uses Jones et al. (2018) implementation of Westfall and Young (1993) correction for multiple hypothesis testing.

* $p < .1$, ** $p < .05$, *** $p < .01$

Next, we note that we have a portion of our sample for whom yield measurements on non-enrolled plots are missing. This is primarily a result of treated farmers harvesting non-enrolled maize without enumerators present to accurately measure yields. We might be concerned that excluding these farmers would bias our treatment effect estimates.

To address this concern, we use two strategies to impute missing yields. First, we compute the average per-acre yields for control farmers within the same field office area, and multiply those by the farmer's non-enrolled acreage to "fill in" per-farmer yields. Second, we instead use the average non-enrolled yields per acre among farmers within the same farmer group. Tables C.3 and C.5 show that the treatment effect estimate on per-farmer yields remains positive and significant, although slightly smaller in magnitude. We expect that the first imputation to be particularly conservative, but in practice the two show very similar results.

Tables C.4 and C.6 similarly show that the estimated treatment effect on profits is smaller in magnitude, but still significant at a 10% level. The decline in magnitude is likely due to the measurement error introduced by simple imputation strategies, since the imputed values are a noisy proxy for the true yields of these missing farmers.

D Assorted tables for pre-exposed sample

Table C.3: Maize yields per farmer with missing yields imputed, primary sample

	(1) Maize Per Farmer	(2) Maize PF + Ctrl Avg	(3) Maize PF + Group Avg
1AF participant	345.074***	224.639***	220.278***
	(84.088)	(68.097)	(68.059)
Married	-41.144	-32.893	-36.291
	(152.252)	(131.275)	(131.274)
Father, 2ary school (0/1)	249.461***	214.598**	213.033**
	(94.245)	(82.203)	(82.123)
Farm labor >50% income	-10.724	29.246	26.358
	(84.821)	(77.787)	(77.869)
Used ag tech 2016	-100.434	-145.934*	-146.689*
	(86.068)	(78.101)	(78.076)
Prev. 1AF knowledge	114.447	145.939	140.021
	(166.821)	(139.467)	(138.341)
Intercropped 2016	-138.885	-78.371	-78.056
	(94.922)	(86.772)	(86.741)
Credit access 2016	-65.529	-35.722	-34.615
	(84.676)	(72.635)	(71.769)
Household size	35.935**	37.132**	37.307**
	(17.200)	(14.402)	(14.414)
FAW Incidence	-1.091	1.551	1.635
	(8.499)	(7.419)	(7.428)
Maize acres, 2016	413.300***	373.693***	374.013***
	(79.716)	(63.613)	(63.596)
Maize yield/acre, 2016	0.461***	0.472***	0.471***
	(0.108)	(0.101)	(0.100)
Observations	637	765	765
R^2	0.282	0.245	0.244
Control Mean Dep. Var	1081.431	1081.431	1081.431

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions which include field office area fixed effects. Ctrl Avg uses field office level averages of control farmer yields per acre to fill in missing non-enrolled yields for treated farmers. Group Avg uses instead farmer group level non-enrolled averages.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table C.4: Profit with missing yields imputed, primary sample

	(1) Profit	(2) Profit + Ctrl Avg	(3) Profit + Group Avg
1AF participant	79.834***	45.964*	44.311*
	(29.996)	(24.491)	(24.467)
Married	-19.811	-16.315	-17.604
	(55.475)	(47.939)	(47.927)
Father, 2ary school (0/1)	85.171**	68.306**	67.713**
	(34.894)	(30.455)	(30.447)
Farm labor >50% income	1.530	17.998	16.903
	(31.330)	(27.940)	(27.956)
Used ag tech 2016	-49.228	-58.869**	-59.155**
	(30.826)	(28.115)	(28.084)
Prev. 1AF knowledge	22.103	50.363	48.119
	(61.844)	(53.984)	(53.319)
Intercropped 2016	-45.234	-26.844	-26.725
	(34.524)	(31.750)	(31.724)
Credit access 2016	-28.493	-19.153	-18.733
	(30.171)	(26.993)	(26.627)
Household size	13.854**	14.140***	14.206***
	(6.035)	(5.061)	(5.066)
FAW Incidence	0.668	1.214	1.246
	(3.177)	(2.767)	(2.771)
Maize acres, 2016	126.476***	113.723***	113.844***
	(28.411)	(22.914)	(22.900)
Maize yield/acre, 2016	0.149***	0.155***	0.155***
	(0.039)	(0.037)	(0.037)
Observations	637	765	765
R^2	0.236	0.202	0.203
Control Mean Dep. Var	337.700	337.700	337.700

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions which include field office area fixed effects. Ctrl Avg uses field office level averages of control farmer yields per acre to fill in missing non-enrolled yields for treated farmers. Group Avg uses instead farmer group level non-enrolled averages.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table C.5: Maize yields per farmer with missing yields imputed, full sample

	(1) Maize Per Farmer	(2) Maize PF + Ctrl Avg	(3) Maize PF + Group Avg
1AF participant	281.794***	177.563***	174.706***
	(49.697)	(41.689)	(41.790)
Married	40.816	44.829	46.202
	(73.664)	(65.763)	(65.430)
Father, 2ary school (0/1)	203.929***	157.650***	156.219***
	(51.499)	(45.452)	(45.422)
Farm labor >50% income	102.909*	109.748**	107.762**
	(55.483)	(49.220)	(49.188)
Used ag tech 2016	-44.779	-72.063	-72.484
	(67.917)	(58.976)	(58.994)
Prev. 1AF knowledge	63.991	98.291*	96.923*
	(55.526)	(52.068)	(51.959)
Intercropped 2016	-52.260	-7.533	-8.042
	(58.359)	(52.627)	(52.584)
Credit access 2016	-27.312	5.179	6.742
	(52.687)	(45.207)	(44.818)
Household size	41.364***	42.865***	42.858***
	(9.691)	(8.449)	(8.435)
FAW Incidence	1.163	2.706	2.629
	(4.320)	(4.041)	(4.031)
Maize acres, 2016	393.505***	381.243***	381.223***
	(56.672)	(49.450)	(49.413)
Maize yield/acre, 2016	0.379***	0.374***	0.374***
	(0.085)	(0.073)	(0.073)
Observations	1785	2137	2137
R^2	0.225	0.207	0.207
Control Mean Dep. Var	1145.997	1145.997	1145.997

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions which include field office area fixed effects. Ctrl Avg uses field office level averages of control farmer yields per acre to fill in missing non-enrolled yields for treated farmers. Group Avg uses instead farmer group level non-enrolled averages.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table C.6: Profit with missing yields imputed, full sample

	(1) Profit	(2) Profit + Ctrl Avg	(3) Profit + Group Avg
1AF participant	52.688***	26.065*	24.982
	(18.379)	(15.617)	(15.636)
Married	3.042	6.267	6.755
	(26.824)	(24.088)	(23.989)
Father, 2ary school (0/1)	68.707***	51.388***	50.847***
	(18.959)	(16.581)	(16.564)
Farm labor >50% income	43.330**	45.323**	44.571**
	(20.864)	(18.168)	(18.153)
Used ag tech 2016	-24.620	-30.513	-30.572
	(25.298)	(21.678)	(21.681)
Prev. 1AF knowledge	14.309	27.096	27.102
	(19.675)	(17.308)	(17.301)
Intercropped 2016	-11.153	4.277	4.034
	(21.411)	(19.276)	(19.260)
Credit access 2016	-20.797	-6.713	-6.098
	(19.151)	(16.703)	(16.535)
Household size	13.645***	14.399***	14.399***
	(3.532)	(3.086)	(3.081)
FAW Incidence	0.784	1.078	1.052
	(1.617)	(1.485)	(1.483)
Maize acres, 2016	123.842***	118.645***	118.630***
	(20.187)	(17.747)	(17.731)
Maize yield/acre, 2016	0.118***	0.116***	0.116***
	(0.030)	(0.026)	(0.026)
Observations	1785	2137	2137
R^2	0.178	0.164	0.164
Control Mean Dep. Var	361.164	361.164	361.164

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions which include field office area fixed effects. Ctrl Avg uses field office level averages of control farmer yields per acre to fill in missing non-enrolled yields for treated farmers. Group Avg uses instead farmer group level non-enrolled averages.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table D.1: Study year (2017) practices, pre-exposed sample

	(1) Imp. Seed	(2) Int. Beans	(3) CAN AP	(4) DAP AP	(5) CAN PP	(6) DAP PP	(7) Used Plow	(8) Row Spacing	(9) Plant Spacing
1AF participant	-0.007	0.022	0.006	0.429***	0.199***	-0.103***	0.066***	0.237***	0.189***
	(0.008)	(0.024)	(0.007)	(0.025)	(0.021)	(0.020)	(0.023)	(0.026)	(0.026)
Married	0.019	-0.051	-0.004	-0.014	0.010	0.007	-0.070***	0.007	0.057
	(0.013)	(0.036)	(0.010)	(0.030)	(0.026)	(0.030)	(0.023)	(0.036)	(0.037)
Father, 2ary school (0/1)	-0.012	0.022	-0.009	0.026	0.023	-0.014	0.041**	0.005	0.012
	(0.009)	(0.024)	(0.006)	(0.018)	(0.016)	(0.018)	(0.019)	(0.028)	(0.022)
Farm labor >50% income	0.022**	0.094***	-0.003	0.007	-0.005	-0.007	-0.073***	0.022	-0.006
	(0.010)	(0.028)	(0.008)	(0.023)	(0.022)	(0.025)	(0.025)	(0.031)	(0.027)
Used ag tech 2016	0.624***	0.008	-0.027*	0.083***	0.062*	-0.006	0.041	-0.005	0.015
	(0.041)	(0.040)	(0.015)	(0.030)	(0.035)	(0.037)	(0.031)	(0.048)	(0.034)
Prev. 1AF knowledge	0.011	0.011	-0.001	0.083***	0.036*	-0.006	0.019	-0.011	-0.005
	(0.012)	(0.030)	(0.006)	(0.022)	(0.021)	(0.025)	(0.020)	(0.033)	(0.027)
Intercropped 2016	0.019**	0.172***	0.006	-0.012	0.018	0.004	0.004	0.038	0.001
	(0.010)	(0.028)	(0.009)	(0.023)	(0.018)	(0.022)	(0.018)	(0.030)	(0.024)
Credit access 2016	-0.005	-0.012	-0.003	-0.027	0.019	0.038	0.010	-0.018	0.010
	(0.010)	(0.030)	(0.007)	(0.026)	(0.022)	(0.025)	(0.026)	(0.027)	(0.026)
Household size	0.004**	0.011**	-0.001	0.000	-0.001	0.001	0.005	0.004	0.001
	(0.002)	(0.005)	(0.001)	(0.003)	(0.004)	(0.005)	(0.004)	(0.006)	(0.005)
FAW Incidence	-0.001	-0.002	-0.000	0.002	0.001	-0.001	0.004*	0.003	-0.003*
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Maize acres, 2016	0.013***	-0.004	0.013**	0.005	0.018	-0.014	0.038***	0.005	0.007
	(0.005)	(0.017)	(0.006)	(0.015)	(0.014)	(0.014)	(0.012)	(0.019)	(0.020)
Maize yield/acre, 2016	0.000***	0.000	0.000	0.000*	0.000	-0.000	0.000**	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	1372	1361	1372	1372	1372	1372	1372	1372	1372
R^2	0.638	0.274	0.020	0.335	0.163	0.148	0.222	0.085	0.091
Ctrl Mean Dep. Var	0.923	0.536	0.011	0.548	0.765	0.249	0.781	0.430	0.142

Standard errors in parentheses

Note: standard errors clustered at farmer group level. Results in this table are from OLS regressions which include field office area fixed effects. Imp. Seed refers to use of commercial improved seed. AP stands for at planting, and PP stands for post planting. 1AF recommends using DAP at planting and CAN post planting. Row spacing and plant spacing are 1 if average spacing within 5 cm of recommended.

* $p < .1$, ** $p < .05$, *** $p < .01$

E 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. Farm labor is valued using a local day wage for agricultural labor, devalued by 50% (roughly the rural unemployment rate according to DHS data). Profit is simply the difference between projected farmer revenues and costs.

F Pre-Analysis Plan Review Letter

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

G More on Separability

Table G.1: Labor use and household size, primary sample

	(1) Log Unpaid Labor	(2) Log Paid Labor	(3) Log Total Labor
1AF participant	0.241***	-0.056	0.201***
	(0.060)	(0.087)	(0.044)
Log HH Size	0.145*	0.018	0.104**
	(0.076)	(0.094)	(0.046)
Log Owned Acres Cultivated	0.186***	0.460***	0.310***
	(0.038)	(0.065)	(0.028)
Log Cluster Median Wage	-0.023	-0.125	-0.030
	(0.167)	(0.259)	(0.118)
Observations	669	466	684
R^2	0.158	0.160	0.294

Standard errors in parentheses

Note: standard errors clustered at cluster level. Results are from OLS regressions which include field office area fixed effects. Labor is measured as person-days used for land prep, planting, and post-planting.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table G.2: Labor use and household size, pre-exposed sample

	(1) Log Unpaid Labor	(2) Log Paid Labor	(3) Log Total Labor
1AF participant	0.291***	0.089	0.228***
	(0.050)	(0.059)	(0.034)
Log HH Size	0.367***	-0.033	0.196***
	(0.052)	(0.071)	(0.033)
Log Owned Acres Cultivated	0.110***	0.310***	0.225***
	(0.034)	(0.039)	(0.024)
Log Cluster Median Wage	-0.378**	0.100	-0.145
	(0.153)	(0.168)	(0.094)
Observations	1188	909	1215
R^2	0.170	0.093	0.240

Standard errors in parentheses

Note: standard errors clustered at cluster level. Results are from OLS regressions which include field office area fixed effects. Labor is measured as person-days used for land prep, planting, and post-planting.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table G.3: Labor use and household size, primary sample

	(1) IHS Unpaid Labor	(2) IHS Paid Labor	(3) ISH Total Labor
1AF participant	0.315***	0.286**	0.200***
	(0.076)	(0.134)	(0.044)
Log HH Size	0.146*	-0.246**	0.104**
	(0.081)	(0.116)	(0.045)
Log Owned Acres Cultivated	0.172***	0.448***	0.309***
	(0.044)	(0.088)	(0.028)
Log Cluster Median Wage	0.077	-0.276	-0.030
	(0.225)	(0.375)	(0.118)
Observations	684	684	684
R^2	0.124	0.084	0.294

Standard errors in parentheses

Note: standard errors clustered at cluster level. Results are from OLS regressions which include field office area fixed effects. Labor is measured as person-days used for land prep, planting, and post-planting.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table G.4: Labor use and household size, pre-exposed sample

	(1) IHS Unpaid Labor	(2) IHS Paid Labor	(3) IHS Total Labor
1AF participant	0.387***	0.208**	0.230***
	(0.056)	(0.092)	(0.033)
Log HH Size	0.457***	-0.250**	0.200***
	(0.067)	(0.102)	(0.034)
Log Owned Acres Cultivated	0.074**	0.372***	0.225***
	(0.037)	(0.062)	(0.024)
Log Cluster Median Wage	-0.431***	0.393	-0.184**
	(0.164)	(0.277)	(0.087)
Observations	1216	1216	1216
R^2	0.175	0.059	0.237

Standard errors in parentheses

Note: standard errors clustered at cluster level. Results are from OLS regressions which include field office area fixed effects. Labor is measured as person-days used for land prep, planting, and post-planting.

* $p < .1$, ** $p < .05$, *** $p < .01$