The effects of relaxing information and credit constraints on productivity: 5-Year experimental evidence from self-reported, GPS-based, and remotely sensed data[1]

Abdulrazzak Tamim[a,*], Aurélie P. Harou[b], Marshall Burke[c,d], David Lobell[c,d], Malgosia Madajewicz[e], Christopher Magomba[f], Hope Michelson[g], Cheryl Palm[h], Jenny Xue[d]

[a]University of California, Berkeley, Center for Effective Global Action
[b]McGill University, Department of Natural Resource Sciences, Agricultural Economics
[c]Stanford University, Department of Earth System Science
[d]Stanford University, Center on Food Security and the Environment
[e]Columbia University, Center for Climate Systems Research
[f]Sokoine University of Agriculture, Department of Agricultural Economics and Agribusiness
[g]University of Illinois at Urbana-Champaign, Department of Agricultural and Consumer Economics
[h]University of Florida, Department of Agricultural and Biological Engineering

Abstract. We examine the longer-term effects of a randomized control trial (RCT) conducted between 2014-2016 on fertilizer adoption and productivity. The initial RCT randomly assigned vouchers (V), plot-specific fertilizer recommendations (R), or both treatments (R+V) to Tanzanian farmers to ease credit and information constraints. The findings of the shorter-term study are that respondents in the V and R+V groups applied more fertilizers, but that resulted in higher productivity for the R+V group only, with an estimated net benefit that is equivalent to wages from 21 days of work. In this study, we follow respondents over an additional three years and report two main results. First, although farmers do not apply fertilizer once vouchers are no longer provided, when we include new untested plots that farmers cultivate over the longer term, respondents in the R+V group still sustain higher yields in 2019 using self-reported data. Second, by using GPS measures of plot size and remote sensing estimates of productivity, we find that none of the treatment effects are robust when using different data sources, with productivity increasing by 0-37% in 2016 and 0-33% in 2019. Our results highlight the importance of extending follow-up periods in RCTs through sustainable and repeated interventions to enhance adoption. The findings also emphasize the role of using more reliable data as analysis and policy implications follow from how outcomes are measured.

Keywords: information, credit, sub-Saharan Africa, fertilizer, soil perceptions.

JEL Codes: O13, Q16, Q18
***Corresponding author:** Abdulrazzak Tamim, abdulrazzaktamim1@gmail.com

1 Introduction

Although many developing countries have witnessed increased productivity since the Green Revolution, staple cereal yields remain low in a number of low-income countries, especially in sub-Saharan Africa (SSA). Numerous factors contribute to this low productivity, including widespread underuse of modern agricultural technologies such as mineral fertilizer, which has the potential to increase profitability as shown in different studies (Duflo, Kremer, & Robinson,

[1]Last updated on December 23, 2020. This research is funded by the SSHRC grant 430-2018-1121 and Stanford University. We thank Matthieu Chemin for providing helpful suggestions. The data used in this paper were collected by Aika Aku, Venacky Daniel, Ambonisye Haule, Litwagile Kajela, Vulfrida Kessy, Sweetness Kileo, Steven Kuhanwa, Habibu Lupatu, Josephat Mbepera, Lilian Mduda, Saidi Mhina, Angel Minja, Paschal Mmelo, Anna Muffui, Burton Mwambokela, Rajabu Mwanyika, and Gloria Peter from Sokoine University of Agriculture under the supervision of Abdulrazzak Tamim, Christopher Magomba, and Revocatus Kayaga Ntengo. The original study was registered in the AEA RCT Registry and the unique identifying number is: "AEARCTR-0005470". This study was approved by the Research Ethics Board at McGill University, file number 16-0619.

2008; Harou, Walker, & Barrett, 2017; Kaizzi et al., 2012; Vanlauwe et al., 2011; Xu, Guan, Jayne, & Black, 2009). Many researchers have studied the lack of fertilizer adoption and documented different explanations for it, including poor soil nutrient content, making the soil irresponsive to fertilizer application, high transaction costs and prices of procuring fertilizer, non-adoption among peers in the same social network, risk and inability to access credit to purchase fertilizer, and absence of information about the suitability of fertilizer types and quantities to apply (Bandiera & Rasul, 2006; Conley & Udry, 2010; Croppenstedt, Demeke, & Meschi, 2003; Duflo, Kremer, & Robinson, 2011; Emerick, de Janvry, Sadoulet, & Dar, 2016; Harou et al., 2020; Jama, Kimani, Harawa, Mavuthu, & Sileshi, 2017; Marenya & Barrett, 2009a, 2009b; Suri, 2011).

In this study, we visit farmers three years after they participated in a randomized control trial (RCT) explained in Harou et al. (2020) that tested the role of credit constraints and the absence of site-specific fertilizer information on fertilizer adoption rates in Tanzania between 2014-2016. The initial study had three treatment arms: farmers who were given plot-specific fertilizer recommendations only (group R); farmers who received vouchers only to purchase any fertilizer they wanted (group V); and farmers who received both plot-specific fertilizer recommendations and vouchers (group R+V). Harou et al. (2020) found that only farmers who were given vouchers increased fertilizer use in 2016 (groups V and R+V), and only those among them who got recommendations (group R+V) witnessed an increased yield. There are two ways to interpret these results. First, farmers cannot afford fertilizer given the available financial services and any biases in their decision making; second, farmers are not sufficiently convinced of the profitability of fertilizer to risk their own resources on it.

The follow-up study may help to distinguish these two explanations. A continued effect on fertilizer use and yields would be evidence of the second interpretation since vouchers were no longer provided after the initial experiment. On the other hand, an absence of a sustained treatment effect is ambiguous since farmers may either not be able to afford fertilizer once they are no longer provided with vouchers, or they may not have been sufficiently convinced of profitability by the results of the first experiment. Given these possible mechanisms that may explain findings from Harou et al. (2020), we examine the longer-term effects of the original intervention in order to shed light on the possible fertilizer adoption constraints. Indeed, few studies we are aware of examine the effects of relaxing farmers' constraints in the longer-run. However, understanding the potential effects of such interventions several years after their implementation is important in thinking about designing policies that are sustainable and can have lasting effects. Hence, we conducted a survey in August 2019 with the same 1,050 households who participated in Harou et al. (2020) asking them about fertilizer application rates and decisions, maize yields, plot characteristics, farmers' perceptions of soil quality, assets, credit, and retention of information. We also collected polygons of farmers' plots using GPS devices, making this study a first systematic longer-term effort that compares productivity using survey data, GPS estimates, and remote sensing techniques.

Our results indicate that, first, using self-reported (SR) measures of productivity and by employing a difference-in-differences (DD) estimation, farmers in the R+V group have a 25% increase in their main maize plot's (MMP) productivity in 2017, but this productivity becomes insignificant in 2018 and 2019. When we include new plots that were not tested as a part of the original RCT in 2014 but that farmers switch to cultivate over 2016-2019, we find a yearly increase in productivity of around 33% from the control group's average over 2016-2019.

Second, since SR plot size is measured with error and can be imprecise while areas measured through Global Positioning Systems (GPS) may be more accurate (Abay, Abate, Barrett, & Bernard, 2019; Abay, Bevis, & Barrett, 2020; Carletto, Gourlay, & Winters, 2015; Carletto, Savastano, & Zezza, 2013), we use GPS measures to estimate the area rather than relying on SR data only. Furthermore, we use unique remote sensing methods explained in Lobell et al. (2019) to estimate productivity derived from satellite images. Our results that rely on endline OLS regressions indicate no or minor treatment impact (0-4%) on productivity when using satellite-based productivity. On the other hand, GPS-based productivity measures do not show an impact in 2016, but the DD estimates indicate a treatment impact of around 30% in 2019, only when new plots that were not tested in the original intervention are included in the analysis. We provide suggestive evidence that this treatment impact may be driven by over-reporting output, which would result in an increase in productivity despite controlling for measurement error in plot size.

Third, we examine fertilizer adoption and find that farmers do not apply fertilizers when vouchers are no longer provided, over 2017-2019. To understand the mechanisms that impede fertilizer adoption, we investigate three possible pathways – first, how much information farmers actually retain in the longer-term; second, whether farmers update their subjective soil beliefs after applying fertilizer; third, their willingness to pay (WTP) for soil information. We find that farmers in the R+V group do not update their soil perceptions, and that they are not willing to pay more for fertilizer compared to the control group. Therefore, farmers seem not to believe that using fertilizer is profitable, perhaps because of the minor treatment impact, which is supported by satellite estimates.

Although the literature on fertilizer adoption in SSA is vast and spans many decades, this research contributes to the literature in three novel ways. First, we extend an existing RCT and measure adoption and productivity over a relatively long period of time with a high tracking rate as we resurvey 93.5% from the sample interviewed in 2016. In doing so, we are able to disentangle the mechanisms that hinder investment in fertilizer – namely, that farmers may not be convinced of the profitability of fertilizer rather than merely facing binding credit constraints. Relative to the broader technology adoption literature, we have rich panel data obtained by tracing farmers back in 2019, five years after the baseline survey was conducted, while the majority of other technology adoption field studies conduct their endline one year since their intervention starts, with some exceptions such as Duflo et al. (2011) who collect data on three seasons in their field experiment that focuses on biases in decision-making and Beaman, BenYishay, Magruder, and Mobarak (2018) who follow farmers for three years to study knowledge and adoption of pit planting and crop residue management.

Second, we use a novel approach to corroborate our findings between SR, GPS-based, and satellite-derived productivity. The initial yield results stem from SR answers, which are prone to recall bias and measurement error. To test the degree to which these biases might affect our results, we use an independent method to measure yields using satellite images (Lobell et al., 2019). This is the first study we are aware of that uses this methodology to corroborate findings from a field experiment. The results of this study highlight the role of using different data sources and empirical strategies in influencing research findings and policy implications, calling for caution when interpreting data that may be measured with error.

Third, we follow a theoretical and an empirical approach to assess if farmers update their soil perceptions after receiving information about their soils' fertility and/or a voucher. While other studies have looked at factors affecting soil perceptions, often finding yields to be a significant determinant of perceived soil fertility, these studies provides insights about correlates of soil perceptions, but the estimates can be biased due to reverse causality and omitted variable bias (Berazneva, McBride, Sheahan, & Güereña, 2018; Marenya, Barrett, & Gulick, 2008). This study is the first to use randomization to understand if farmers update their prior beliefs about the soil fertility after applying fertilizer, which may help in explaining the lack of adoption through farmers' incorrect soil perceptions, highlighting the importance of policies that target farmers' beliefs.

The remainder of this paper is organized as follows: section 2 provides some background about the study area. In section 3, we detail our research design and data sources. Section 4 presents the econometric specifications, and section 5 presents the results. Then, we discuss our results in section 6, and section 7 concludes.

2 Background

This study takes place in Morogoro Rural, one of the six districts in Tanzania's Morogoro Region (Figure 1). Maize is an important income source for many of the region's smallholder farmers and is also important for consumption, constituting 60 percent of dietary calories in Tanzania (Mtaki, 2017). However, maize yields have been low in the region, with average yields of 1905 kilograms per hectare in 1994-1995 compared to 2000 kilograms per hectare in 2000-2001 (Paavola, 2008). Most households do not use improved seeds or fertilizers. Harou et al. (2020) find that 20 percent of Tanzanian farmers in their sample used improved maize seeds in 2014, while less than one percent adopted mineral fertilizers. Overall, rural households are worse off than their urban counterparts as they have much lower income and consumption levels, higher poverty incidence, and less access to markets (Paavola, 2008).

In 1993, the government of Tanzania revised its 1984 fertilizer recommendations by conducting trials that aimed to study the crop response of fertilizer applications. The 1993 study refined the previous recommendations of 63 agro-ecological zones. Recommendations in the Morogoro Region vary between 40-90 kilograms of nitrogen and 17-20 kilograms of phosphorus for one hectare of maize depending on the agro-ecological zone (Mowo, Floor, Kaihura, & Magoggo, 1993). For example, in the Morogoro Region, the government recommends supplying the soil with 40 and 20 kilograms of nitrogen and phosphorus per one hectare of maize, respectively, in a site called Ilonga, but 90 and 20, respectively, in another site in Morogoro called Ifakara. The recommendations include only nitrogen and phosphorus, ignoring other nutrients important for soil fertility, such as potassium, sulfur and active carbon. Furthermore, because the soil nutrient limitations in many countries of sub-Saharan Africa (SSA) vary spatially (Rowe, van Wijk, de Ridder, & Giller, 2006; Tittonell et al., 2013; Zingore & Johnston, 2013), these recommendations, are unlikely to be profit-maximizing for all farmers because soils respond in different ways to fertilizer application. Indeed, Harou et al. (2020) find that sulfur is a limiting nutrient in many of the plots they tested - a nutrient not addressed by the government's recommendations.

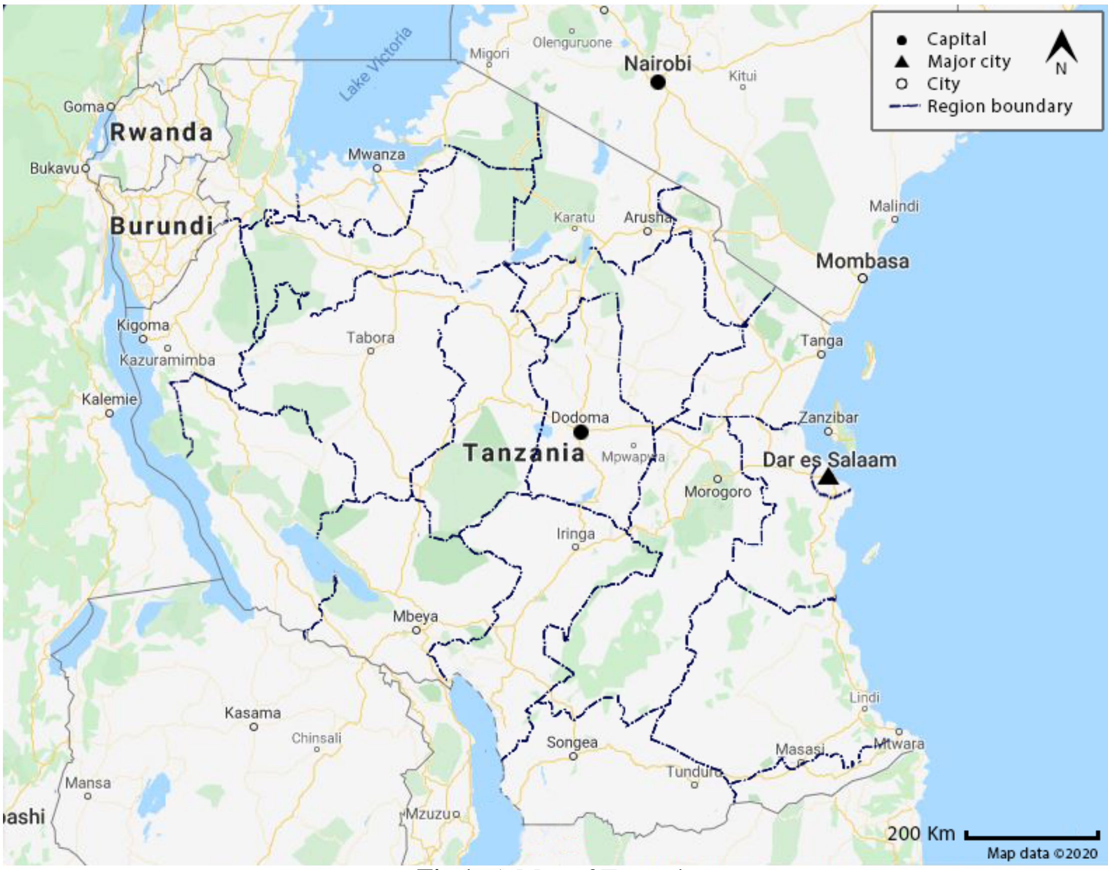

Fig 1: A Map of Tanzania

3 Research Design

3.1 The 2014 Randomized Experiment

Before proceeding to detailing the current design, we provide information about the 2014 RCT. Farmers were randomly invited to participate in the original study in two stages. First, out of all maize-growing and accessible villages in Morogoro Rural, 27 villages were assigned randomly as control and 20 as treatment. Second, within the 20 treatment villages, farmers were randomly allocated to one of four groups (recommendations, vouchers, both treatments, and control). The study was designed this way to allow for the testing of spillovers by comparing control farmers in control villages and control farmers in treatment villages, as discussed in section 3.4.

The number of participants per treatment is shown in Figure 2 (source: (Harou et al., 2020)). We next share some details about the treatment arms; for more detailed information on the original study's design, consult Harou et al. (2020). From each control village, 10 eligible farmers were randomly selected to participate in the study. In treatment villages, the farmers were assigned randomly to one of the following arms (10 farmers per arm):

i. **Plot-specific recommendations (Group R):** Farmers received information about which fertilizer types and quantities they should apply on their 2014 MMP per an acre and half an acre of land area planted with maize. These recommendations were based on soil samples collected and tested by a team of agronomists and soil scientists

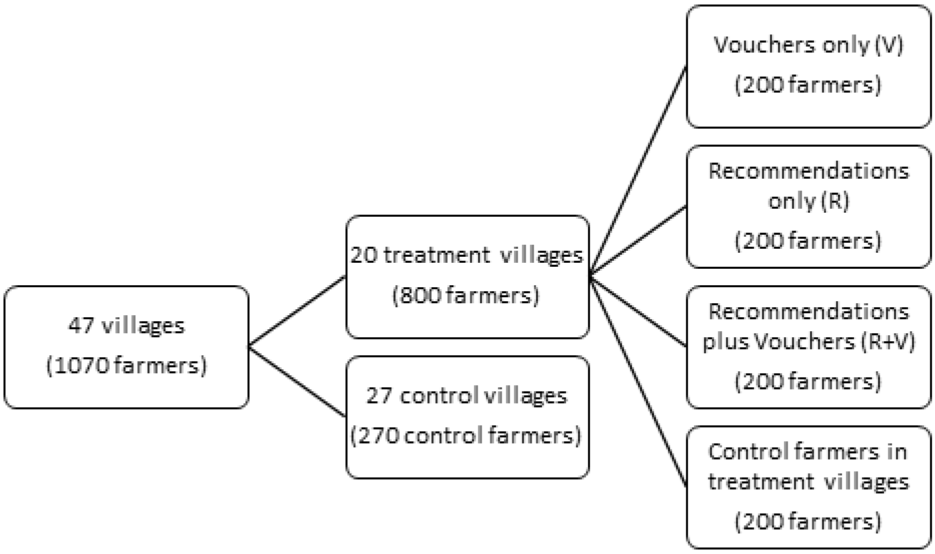

Fig 2: Study Design

from SUA. Agronomists then met with treated farmers and presented them with a card/sheet explaining their soils' deficiencies and what fertilizers were recommended for their MMP.

ii. **Vouchers (Group V):** Farmers in this group were given a voucher valued at 80,000 TZ Shillings (or about 40 USD at the time of the study) that they could redeem to purchase any fertilizer they wanted from a specific agro-input dealer. Farmers in this group also received their plot-specific fertilizer recommendations for the 2016-2017 planting season, after the endline data were collected in 2016.

iii. **Plot-specific recommendations and vouchers (Group R+V):** This group received both the vouchers and recommendations as previously described (treatments R and V).

iv. **Control farmers (Group C):** Control farmers received neither recommendations nor vouchers by 2016. This group consists of control farmers in both control and treatment villages since we do not find an evidence of spillovers, as explained in section 3.4. After the endline in 2016, farmers in the control group were provided with fertilizer recommendations for their 2014 MMP.

3.2 Data Collection

We are interested in assessing the longer-term effects of alleviating credit and information constraints on fertilizer adoption and maize yields using different measures of productivity. In 2019, we resurveyed a group of 1,050 households who participated in the original 2014-2016 RCT. With the collaboration of Sokoine University of Agriculture (SUA), a leading university in Tanzania, we collected primary data from the same villages and farmers in the Morogoro Region which were selected randomly at baseline. The questionnaires included modules on assets, credit, land tenure, maize yields, mineral fertilizer use, questions about farmers' retention of the fertilizer recommendations, and characteristics of each year's MMP, which is defined as the plot that is cultivated with maize and that is the most important plot for the household in terms of food security and income generation. We also collected recall data on some of these same topics for 2017 and 2018, creating a panel dataset of the same households for five years (2014 and 2016-2019). To remind farmers of the 2014 MMP from which soil samples were taken, we showed them a map of their plots and their house location drawn by enumerators in 2014. We also told them how they referred to that plot. Because the original project was highly respected by extension agents and farmers for its impact and from SUA's reputation, we did not have many difficulties in reaching the same farmers. We were able to visit 920 households out of the 1,050 who participated in the original study by Harou et al. (2020), resulting in an overall high tracking rate of around 88% since baseline in 2014 – or 93.5% compared to the 984 farmers surveyed in 2016 – despite a lack of communication with respondents since the initial study was concluded in 2016.

We also collected GPS polygons in 2019, after enumerators were trained and during the survey's two-day field testing. Before proceeding to take the GPS polygons and as a part of the survey, a script reminded enumerators about a protocol developed by our team to be used in order to collect these polygons using GPS devices.[2] Enumerators collected GPS coordinates for farmers' 2014 MMP, 2016 MMP and 2019 MMP. They also asked if farmers cultivated the same plot in 2014 (2016) as their 2019 MMP. If not, the enumerator asked to visit the 2014 (2016) MMP to collect those new GPS coordinates. All polygons were collected after completing the household survey questionnaire. In doing so, we are able to minimize any impact of GPS measures of areas in influencing SR answers since farmers might have seen the GPS estimation results from the handheld devices but only after they reported their areas. Furthermore, we gave enumerators careful instructions about measuring boundaries of the area planted with maize only and not the entire plot so that we map the same plot boundaries using SR ad GPS measures. In addition to collecting the GPS coordinates, the enumerators administered a short questionnaire, including details on whether the plot was intercropped, if there are any large objects such as buildings or large trees inside the perimeter of the plot, and the clarity of the plot's boundaries.

3.3 Outcome Variables of Interest

Our main outcomes are fertilizer adoption, measured in SR kilograms of any fertilizer applied per SR acre of maize planted (kg/acre) for the years 2016-2019, and as a binary variable that equals one if a farmer applied any fertilizer and zero otherwise, and agricultural productivity measured in kilograms of maize harvested per acre of maize planted (kg/acre) during the long rains season of the years 2016-2019. Productivity is measured using SR output divided by SR area, SR output divided by GPS-based area, and using satellite images, explained further in section 5.3.

To explore the patterns behind farmers' adoption decisions, we first assess retention of information about the recommended fertilizer types and quantities in the years 2016 and 2019 as measured through a standardized retention index, with larger values indicating higher retention abilities. Second, we study subjective soil beliefs (SSB), measured by asking farmers "On a scale of 1 to 5, 1 being very poor and 5 being very good, how would you rate the quality of the soil of this MAIZE plot in the year XXXX?" Then, we group SSB to zero (poor) if SSB is equal to one or two, one (fair) if SSB is equal to three, and two (good) if SSB is equal to four or five to follow the way SSB is defined in the literature.[3]

[2] All polygons were collected using Garmin's GPSMAP 64 devices.

[3] The results that we will show later are robust when using the 5-scale measure of perceptions.

We also examine SSB conditional on the true soil quality, which is proxied by a soil index that we create from soil fertility indicators (pH, electrical conductivity, active carbon, nitrogen, sulfur, potassium, and phosphorus). Based on the index, soils are classified as being of poor, fair, or good quality. This procedure results in one unconditional SSB regression, and three regressions conditional on the poor, fair, or good SSB. Third, we study the treatments' effect on WTP for soil information in 2016, measured both as an indicator variable that equals unity if a farmer is willing to pay any amount for soil information, and also in in TZ Shillings elicited through contingent valuation. Enumerators asked farmers if they are willing to pay 1000 TZ shillings to get plot-specific soil recommendations; if a respondent answered "yes", the amount was increased by 500 TZ shillings until the respondent indicated that he/she is not willing to pay the amount, with a maximum possible WTP of 8000 TZ shillings. Although contingent valuation methods may suffer from some bias, our procedure allows for realistic numbers that increase progressively, after some farmers were actually given these recommendations (groups R and R+V), potentially reducing hypothetical bias. Furthermore, compared to the standard contingent valuation method, this procedure is more efficient (Hanemann, Loomis, & Kanninen, 1991).

3.4 Spillovers

A concern about spillovers might arise from farmers sharing their fertilizer recommendations with their neighbors from group V or the control group who were not given such recommendations before endline in 2016. Indeed, a large body of literature investigates how farmers disseminate information in their social networks. Since farmers in group V and the control group were provided with recommendations after endline in 2016, spillovers might have taken place before 2016 only. However, as demonstrated in Harou et al. (2020), no spillovers were detected when comparing control farmers in treatment villages with control farmers in control villages. As a result, in the following analysis, we pool all control farmers together, i.e., we do not differentiate control farmers in treatment villages from control farmers in control villages.

3.5 Attrition

From the 1,050 farmers who participated in the study at baseline in 2014, we were able to revisit 920 farmers in 2019, resulting in an attrition rate of 12.4%, which is relatively low given that five (three) years had passed since the baseline (endline) data collection for Harou et al. (2020). The main reasons for not being able to locate farmers were migration (38.8%) and death (15.5%). Attrition by treatment and control groups is displayed in Table 1. To test whether attrition may have differentially impacted the treatment groups thereby introducing bias, we analyze attrition in Appendix A.1. The analysis indicates that attrition is unlikely to bias our estimates since the probability of attrition does not differ significantly between all treatments and the control group, and attrition is not associated with any of our outcome variables. Because attrition is balanced, we do not adjust for it.

3.6 Baseline Balance

To check if despite randomization there are any baseline differences between the treatments and the control group on outcome variables of interest, as defined earlier, and on control variables, we regress baseline outcome and control

Table 1: Participants by Assinged Group

	Baseline (2014)	First endline (2016)		Second endline (2019)	
	N	N	attrition	N	attrition
V	198	187	5.5%	178	10.1%
R	191	177	7.3%	162	15.2%
R+V	203	190	6.4%	175	13.8%
C	458	430	6.1%	405	11.6%
Total	1050	984	6.3%	920	12.4%

Note: V denotes Voucher group, R denotes the Recommendations group, R+V denotes the Recommendations and Voucher group, and C denotes the Control group.

Table 2: Balance of Outcome Variables

	Fertilizer (kg/acre)	Fertilizer (=1)	Yields (kg/SR acre)	Yields (kg/GPS acre)	SSB=0 poor; 1 fair; 2 good
V	0.44	0.00	-7.87	-35.42	-0.02
	(0.31)	(0.01)	(33.87)	(39.90)	(0.06)
R	0.25	0.01	31.25	-22.91	0.05
	(0.26)	(0.01)	(31.09)	(53.96)	(0.07)
R+V	0.01	-0.00	-22.57	-31.23	-0.01
	(0.02)	(0.01)	(27.51)	(59.97)	(0.05)
N	1,050	1,050	915	752	1,046
R-squared	0.004	0.001	0.003	0.004	0.001
Village FE	YES	YES	YES	YES	YES

Notes: V denotes Voucher group, R denotes the Recommendations group, and R+V denotes the Recommendations and Voucher group. Robust standard errors in parentheses. Standard errors are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

variables on treatment indicators using ordinary least squares (OLS) with village fixed effects (FE) according to the following specification:

$$b_{iv} = \alpha_0 + \sum_{k=1}^{3} \theta_k TREAT_i^k + d_v + \varepsilon_{iv} \tag{1}$$

where b_{iv} is a baseline variable for farmer i in village v, α_0 is a constant, $TREAT_i^k$ is a binary variable that takes the value one for each farmer i assigned to one of the k treatment arms (V, R and R+V) and zero otherwise, d_v is village FE to control for village-specific factors that do not vary over time, and ε_{iv} is the associated idiosyncratic error term. The omitted category is the control group, and standard errors are clustered at the village-level. Table 2 reports the regression results for our main outcomes variables of interest, which shows that there is no significant imbalance at baseline among all outcome variables when comparing the three treatments to the control group. However, an imbalance in SR yields exists when we do not include village FE. Therefore, we check the robustness of our DD results by employing an analysis of covariance (ANCOVA) estimation, as will be detailed in section 4.

Table 3 reports the balance for control variables, all of which defined in Table 21 of Appendix A.2. Most control variables are fairly balanced except for credit access, which is imbalanced between the R group and the control group, and the R+V group and the control group. Since the sample is well-balanced in general, we do not include covariates in all of the econometric specifications.

Table 3: Balance of Covariates

	Dependency	Assets	Age	Gender	Educ	Educ2	Credit	Distance	Total area	Maize Area	Seeds
V	-0.15	-0.00	1.57	-0.03	0.03	0.03	-0.02	0.95	-0.38	0.03	-0.00
	(0.15)	(0.01)	(0.97)	(0.03)	(0.03)	(0.03)	(0.02)	(2.49)	(0.60)	(0.18)	(0.03)
R	-0.05	0.00	2.18*	-0.01	-0.01	-0.00	-0.06**	-5.16	-0.21	0.16	0.00
	(0.13)	(0.01)	(1.22)	(0.03)	(0.04)	(0.02)	(0.03)	(3.75)	(0.68)	(0.23)	(0.03)
R+V	-0.11	-0.00	0.76	0.06*	0.02	0.02	-0.08***	-0.47	-0.74	-0.15	0.02
	(0.11)	(0.01)	(0.58)	(0.03)	(0.03)	(0.02)	(0.03)	(3.35)	(0.72)	(0.14)	(0.04)
N	989	1,050	1,050	1,050	1,050	1,050	1,050	926	1,050	1,046	1,050
R-squared	0.002	0.001	0.003	0.006	0.002	0.002	0.009	0.005	0.002	0.003	0.000
Village FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: V denotes Voucher group, R denotes the Recommendations group, and R+V denotes the Recommendations and Voucher group. All dependent variables are defined in Table 21. Robust standard errors in parentheses. Standard errors are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Compliance with the Assignment to Treatment

	Assigned treatment	Received vouchers N	%	Received recommendations N	%	Meeting attendance N	%
V	187	167	89.3	27	14.4	158	84.5
R	177	5	2.8	128	72.3	121	68.4
R+V	190	175	92.1	168	88.4	175	92.1
C	430	15	3.5	17	4.0	152	35.4

Note: V denotes Voucher group, R denotes the Recommendations group, R+V denotes the Recommendations and Voucher group, and C denotes the Control group.

3.7 Compliance

Some farmers who were randomly assigned to the V, R and R+V treatment arms did not comply with the treatment, as seen in Table 4 showing the numbers and percentages of farmers who received vouchers or recommendations, and farmers who attended a SUA information and/or voucher distribution meeting. The data come from farmer-reported answers to the endline survey that took place in 2016 with a sample size of 984 farmers.

Most farmers who were randomly chosen to receive vouchers received them (89.3% and 91.2% of farmers in groups V and R+V, respectively). Surprisingly, a small number of farmers, 2.8% and 3.5% of farmers in the R and C groups, respectively, report receiving vouchers. This might have occurred because some households who were given vouchers exchanged them with others. As for the recommendations, 88.4% of farmers in the R+V group and 72.3% of farmers in the R group report receiving them. Although all farmers in the V and C groups were supposed to receive the recommendations only after 2016, 14.4% and 4.0% of them, respectively, indicated receiving recommendations by 2016. This might be explained by farmers confusing receiving recommendations from agronomists performing soil testing on their MMP, since soil testing was performed on all farmers' MMP after baseline.

When asked if they attended a meeting in which vouchers and/or fertilizer recommendations cards/sheets were distributed, 84.5%, 68.4%, and 92.1% of farmers in the V, R, and R+V treatments respectively confirmed that they attended (Table 4). In general, there is partial compliance with the treatments. To account for non-compliance, we follow an intent-to-treat analysis, regardless of whether or not farmers received the treatment in the field. This approach deals with partial compliance but leads to an underestimation of the treatment effects.

4 Econometric Specifications

We are interested in measuring the longer-term effects of the R, V, and R+V treatments on fertilizer adoption and maize yields over 2016-2019. To do so, we follow an intent-to-treat (ITT) analysis and estimate the following village fixed effects (FE) model using ordinary least squares (OLS):

$$Y_{ivt} = \alpha_0 + \sum_{k=1}^{3} \theta_k TREAT_i^k + \sum_{t=2016}^{2019} \beta_t d_t + \sum_{t=2016}^{2019} \sum_{k=1}^{3} \gamma_{kt} TREAT_i^k \times d_t + d_v + \varepsilon_{ivt} \tag{2}$$

where Y_{ivt} is an outcome variable for farmer i in village v at time t (2016-2019), α_0 is a constant, $TREAT_i^k$ are dummy variables for the three treatment arms, R, V, and R+V, d_t are year FE to control for aggregate trends that might influence farmers' adoption and productivity, the d_v's are village FE to control for village-specific factors that do not vary over time, and ε_{ivt} is the associated idiosyncratic error term that varies across individuals, between villages and over time. The main coefficient of interest is γ_{kt}, which is the difference-in-differences (DD) estimator for treatment k (V, R or R+V) at time t (2016-2019). In order to estimate the treatment effect in the post-treatment period, we omit the indicator for baseline (2014) among the years' indicators (d_t). Among groups interacted with years, we omit the control group interacted with all years since they are perfectly collinear with the year FE, and the year 2014 interacted with all groups since they are perfectly collinear with the village FE. We cluster standard errors at the village level to account for potential within-village correlation.

Our key identifying assumption is the parallel trend assumption, which indicates that in the absence of the treatment, averages of the control and treatment groups' outcomes follow a parallel path. Given that the treatments were assigned randomly, parallel trends should hold. However, because SR yields are imbalanced when we do not control for village-specific heterogeneity, we explore the robustness of our results in section 5.6 by employing an ANCOVA estimation with the following specification:

$$Y_{ivt} = \alpha_1 + \varphi Y_{i2014} + \sum_{k=1}^{3} \gamma_k TREAT_i^k + \sum_{t=2016}^{2019} \beta_t d_t + M_{i2014} + d_v + \varepsilon_{ivt} \tag{3}$$

where Y_{iv2014} is an outcome measured at baseline, and the remaining variables and parameters are the same as was described in (2). To maximize power, we follow the same procedure in Haushofer and Shapiro (2016) by coding missing baseline values as zeros and adding M_{i2014}, an indicator that equals unity if a baseline observation is missing.[4]

In the presence of a positive treatment effect, Angrist and Pischke (2008) derive analytically a bracketing relationship between DD and ANCOVA – namely, DD overestimates the treatment impact if strict exogeneity holds, whereas ANCOVA underestimates the effect if parallel trends holds. As McKenzie (2012) indicates, the DD estimate is $\gamma_{kt}^{DD} = (\bar{Y}_t^K - \bar{Y}_t^C) - (\bar{Y}_{2014}^K - \bar{Y}_{2014}^C)$, while the ANCOVA estimate is $\gamma_{kt}^{ANCOVA} = (\bar{Y}_t^K - \bar{Y}_t^C) - \hat{\varphi}(\bar{Y}_{2014}^K - \bar{Y}_{2014}^C)$.

[4]Overall, this procedure results in recovering 103 missing observations for the 2014 MMP's SR productivity in 2016, 68 observations in 2017, 55 observations in 2018, and 37 observations in 2019. As for the 2016/19 MMP, the analogous numbers are 112 in 2016, 87 in 2017, 83 in 2018, and 80 in 2019. When productivity is measured using GPS areas, the missing values recovered are 86 in 2016 and 39 in 2019 for the 2014 MMP, and 94 in 2016 and 103 in 2019 for the 2016/19 MMP.

Thus, we can think of the DD estimate as an upper bound of the treatment impact and the ANCOVA estimate as a lower bound.

We also estimate the treatment impact using satellite images of farmers' MMPs. For each polygon, we extract the average value for all available dates for the visible and near-infrared (NIR) bands of Sentinel-2 satellite measurements, which have 10m spatial resolution. From these we compute for each date the Green Chlorophyll Vegetation Index (GCVI), which is defined as:

$$GCVI = \frac{NIR}{(Green - 1)} \tag{4}$$

From the time series of GCVI values for each field, we fit a recursive harmonic regression to make the data more robust to missing observations during the peak of the growing season. Then using the 10th iteration of the harmonic curve we extract the maximum GCVI value over the growing season, which has been shown to correlate well with crop yields in the region (Jin et al., 2019). Because reliable satellite images are only available starting in 2016, we are unable to predict yields via satellite at baseline. We therefore have to rely on OLS to estimate the treatment impact on satellite-derived productivity measures instead of using DD or ANCOVA. We estimate the following OLS model using post-intervention data only:

$$Y_{ivt} = \alpha_2 + \sum_{k=1}^{3} \beta_k TREAT_i^k + d_v + \varepsilon_{ivt} \tag{5}$$

where Y_{ivt} is an outcome of farmer i in village v at post-intervention time t (2016 or 2019), and the remaining variables are the same as was defined in equation (1).

Finally, we use pooled OLS regressions to investigate the correlation between GPS and SR areas by estimating the following equation:

$$GPS_{it} = \alpha_3 + \beta_1 SR_{it} + \varepsilon_{it} \tag{6}$$

where GPS_{it} is the area of the MMP cultivated by farmer i at time t (2014, 2016, and 2019) obtained from GPS estimates, α_0 is again a constant, SR_{it} is the SR area, and ε_{it} is the associated error term. In this estimation, if the slope is close to one and the constant is close to zero, then SR and GPS estimates are very similar.

5 Empirical Results

5.1 Summary Statistics

We start by reporting descriptive statistics of interest before moving onto the estimation strategies and empirical results in section 5.2. Table 5 shows summary statistics for a few variables that describe households at baseline in 2014. Around 10 percent of farmers were able to access credit[5] from any source (banks, cooperatives, family members, etc.) over the last 12 months preceding the baseline survey. While households own on average more than five acres of land, around two acres are planted with maize. Even though 88.8 percent of household heads completed some education, only 5.8 percent completed a level beyond the seven primary Tanzanian levels.

[5] Credit is defined as loans, money borrowed, in-kind credit, and/or money received as a gift.

Table 5: Descriptive Statistics of Baseline Variables

	Mean	SD	Min	Max	N
Panel A: Household characteristics					
Dependency ratio	1.585	1.272	0	8	989
Accessed credit (=1)	0.098	0.298	0	1	1050
Total land area owned (acres)	5.215	6.131	0	50	1050
Area planted with maize (acres)	1.881	1.951	0.07	22.69	1024
Distance to plot (minutes)	34.135	32.314	0	360	926
Panel B: Household head characteristics					
Male (=1)	0.832	0.374	0	1	1050
Age	46.27	14.61	21	87	1050
Education (=1 for any level completed)	0.888	0.316	0	1	1050
Education (=1 if more than primary)	0.0581	0.234	0	1	1050

In Table 6, we present farmers' cultivation behavior. The number of farmers who cultivated their 2014 MMP (from which soil samples were taken in 2014) over 2016-2019 decreased from 822 in 2016 to reach 416 in 2019. Interestingly, more farmers switched to cultivate new plots, with an increase from 81 in 2016 to 369 in 2019 (Table 6). Since some farmers switched their MMP between years, we refer to the plots that farmers consider to be most important for their household in terms of food security and income generation over 2016-2019 as 2016-2019 MMPs, regardless of whether or not soil samples were taken from these plots.

Table 6: Farmers' Cultivation Behavior

	2014	2016	2017	2018	2019	Total
Cultivated the 2014 MMP	915	822	621	499	416	3,273
Regarded the same 2014 MMP as their MMP over 2016-2019	915	807	580	459	338	
Cultivated a new MMP over 2016-2019	0	81	202	251	369	
Total	915	888	782	710	707	4,002

We present summary statistics of fertilizer and yields by year in Table 7. Yields obtained from the 2014 and 2016-2019 MMP do not follow a consistent pattern as they declined between 2014 and 2016, increased between 2016 and 2017, and then declined again over 2017-2019. Fertilizer was applied by a small minority of farmers over the years, increasing from around 1% of farmers in our sample at baseline applying fertilizer on their 2014 MMP to reach 5.2% in 2017, and then dropping again to 2.2% in 2018 and 1.4% in 2019. The year 2016 was an exception as farmers were able to purchase fertilizer because of the vouchers that they were provided with from the original RCT, leading to a quarter of farmers in our sample applying fertilizer.

In Figures 3 and 4, we show histograms of satellite-based yields measured through the GCVI for the 2014 MMP and 2016/19 MMP, respectively. Productivity in 2019 measured through the GCVI is higher than that in 2016 for all plots, which is in accordance with SR data reported in Table 7. In Figures 5 and 6, we compare SR and GPS areas for the 2014 MMP and 2016/19 MMPs, respectively. In both figures, it is evident that farmers over-report the true size of smaller plots and under-report the true size of larger plots, indicating that measurement error, defined as the a difference between SR and GPS areas, declines with GPS plot size, which is consistent with other studies (Abay et al., 2020; Carletto et al., 2015).

Table 7: Descriptive Statistics of Self-Reported Adoption and Productivity

	2014 MMP			2016-19 MMP		
	Fertilizer (kg/acre)	Fertilizer (=1)	Yields (kg/acre)	Fertilizer (kg/acre)	Fertilizer (=1)	Yields (kg/acre)
2014	0.15	0.01	412.16	0.15	0.01	412.16
	(2.52)	(0.09)	(348.87)	(2.52)	(0.09)	(348.87)
2016	7.63	0.25	306.18	7.97	0.25	309.24
	(19.97)	(0.43)	(382.72)	(20.26)	(0.43)	(379.86)
2017	1.52	0.05	502.70	1.48	0.05	513.23
	(8.22)	(0.22)	(401.97)	(8.12)	(0.22)	(414.24)
2018	0.49	0.02	447.07	0.76	0.02	465.92
	(5.11)	(0.15)	(400.23)	(6.83)	(0.15)	(402.64)
2019	0.71	0.01	318.33	0.67	0.02	331.05
	(6.75)	(0.12)	(367.89)	(6.30)	(0.12)	(355.87)

Notes: Numbers are averages. Standard deviations between parentheses.

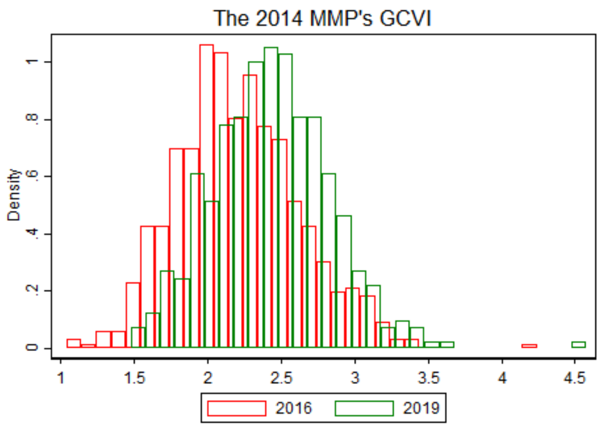

Fig 3: Distribution of the 2014 MMP's GCVI

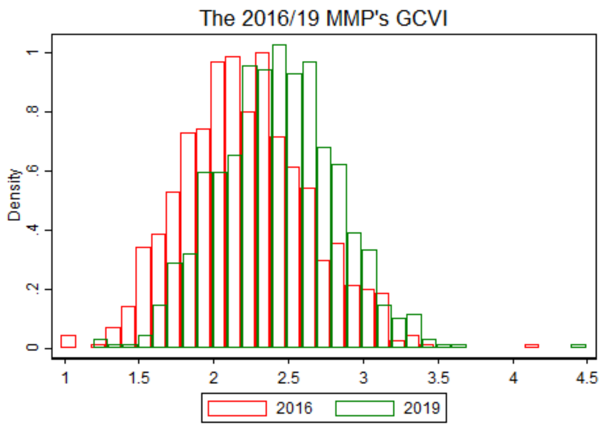

Fig 4: Distribution of the 2016/19 MMP's GCVI

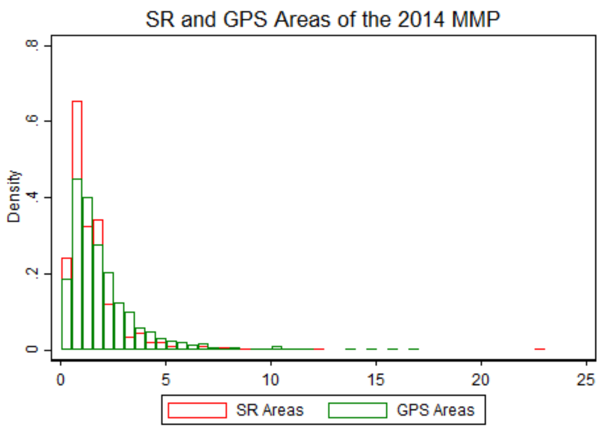

Fig 5: The 2014 MMPs' SR and GPS Size

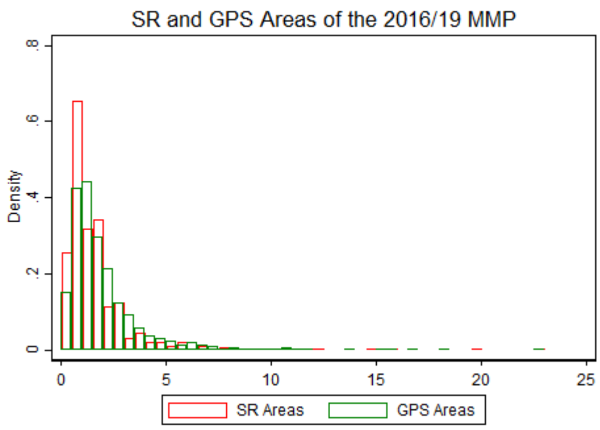

Fig 6: The 2016/19 MMPs' SR and GPS Size

Finally, we show in Figure 7 kernel densities of WTP for plot-specific soil recommendations, classified by four crops (maize, sesame, rice, and cassava). Farmers' WTP increases between 0 and 1000 TZ shillings, but then declines until it reaches around 6000 TZ Shillings, and then increases a bit at about 8000 TZ Shillings, which represents the highest possible amount in the contingent valuation questionnaire. For all WTP of more than 2000 TZ Shillings, farmers are willing to pay most for maize, followed by rice, the most intercropped crop with maize in our sample, and then sesame and cassava, which has the lowest WTP.

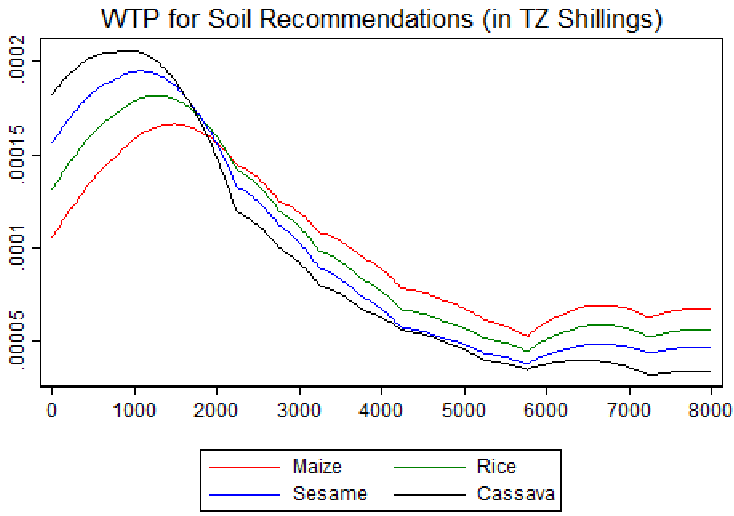

Fig 7: Kernel Densities for WTP in TZ Shillings

5.2 Self-Reported Fertilizer Adoption and Productivity

We start by reporting results from farmers' self-reported (SR) estimations in Table 8, which shows the results of estimating equation (2). The first three columns in Table 8 display the 2014 MMP's results, while the last three columns display the results of the 2016-2019 MMPs. The DD coefficients for fertilizer applied in kilograms per an acre of maize and for the decision to use fertilizer (modelled using a linear probability model) are positive and significant at the 1% level for groups V and R+V in 2016 for both the 2014 MMP and 2016 MMP, but this additional fertilizer translated to higher yields only for the R+V group with a magnitude of 167 kg/acre derived from the 2014 MMP and 152 kg/acre obtained from the 2016 MMP.

When looking at the longer-term treatment impact, we see that farmers in the R+V group in 2017 had a 5% (significant at the 10% level) and a 6% (significant at the 5% level) higher probability of applying fertilizer on their 2014 and 2017 MMP, respectively. The same R+V farmers also obtained an additional 114 kg/acre from the 2014 MMP (column 3 of Table 8) and 139 kg/acre from the 2017 MMP (column 6 of Table 8), both of which are distinguishable from zero at the 5% and 1% levels, respectively. In 2018 and 2019, the DD estimates for fertilizer use are generally negative but small in magnitude, indicating that fertilizer use is back to baseline levels.

Yields obtained from the 2014 MMP are insignificant in 2018 and 2019; however, the 2018 and 2019 MMP yields are significant at the 1% level for the R+V farmers, with additional amounts of 144 and 150 kg/acre, respectively.

Table 8: Treatment Effect on Self-Reported Adoption and Productivity

	2014 MMP			2016-19 MMP		
	Fertilizer (kg/acre)	Fertilizer (=1)	Yields (kg/acre)	Fertilizer (kg/acre)	Fertilizer (=1)	Yields (kg/acre)
2016	0.58**	0.03**	-153.48***	0.88***	0.03***	-147.65***
	(0.22)	(0.01)	(36.57)	(0.32)	(0.01)	(35.75)
2017	1.11***	0.03**	57.73**	0.89***	0.03**	54.01**
	(0.40)	(0.01)	(27.56)	(0.32)	(0.01)	(25.59)
2018	0.46**	0.02*	11.90	0.31**	0.01	12.88
	(0.18)	(0.01)	(35.61)	(0.13)	(0.01)	(29.11)
2019	0.68*	0.01	-132.67***	0.47*	0.01	-136.37***
	(0.39)	(0.01)	(34.67)	(0.25)	(0.01)	(30.91)
V	0.25	0.00	-32.33	0.34	0.00	-42.85
	(0.37)	(0.01)	(37.85)	(0.36)	(0.01)	(36.79)
R	0.06	0.00	7.89	0.14	0.00	-3.16
	(0.32)	(0.01)	(35.32)	(0.31)	(0.01)	(34.30)
R+V	-0.19	-0.00	-47.89	-0.11	-0.00	-58.75**
	(0.18)	(0.01)	(29.29)	(0.16)	(0.01)	(28.63)
2016×V	9.20***	0.31***	80.05	9.17***	0.31***	75.19
	(2.40)	(0.06)	(59.01)	(2.33)	(0.06)	(55.90)
2016×R	1.58	0.04*	16.51	1.26	0.04	11.52
	(1.06)	(0.02)	(49.68)	(1.04)	(0.02)	(45.97)
2016×RV	25.37***	0.72***	166.78***	25.23***	0.72***	152.00***
	(2.28)	(0.03)	(54.91)	(2.15)	(0.03)	(49.65)
2017×V	1.29	0.03	59.89	1.59	0.04	97.84**
	(1.20)	(0.03)	(48.14)	(1.12)	(0.03)	(40.45)
2017×R	-1.12*	-0.04*	-15.10	-0.63	-0.02	4.54
	(0.58)	(0.02)	(55.21)	(0.50)	(0.02)	(47.74)
2017×RV	1.07	0.05*	113.53**	1.34*	0.06**	139.02***
	(0.80)	(0.03)	(51.77)	(0.78)	(0.03)	(48.66)
2018×V	0.76	0.02	93.04	1.00	0.03	96.16
	(1.13)	(0.02)	(61.70)	(1.06)	(0.02)	(59.93)
2018×R	-0.78**	-0.03**	-61.79	-0.17	-0.01	-28.55
	(0.37)	(0.01)	(67.68)	(0.51)	(0.01)	(56.83)
2018×RV	-0.52**	-0.00	72.46	0.95	0.02	143.98***
	(0.20)	(0.02)	(54.38)	(1.06)	(0.02)	(47.44)
2019×V	0.73	0.01	73.41	0.88	0.01	98.81*
	(1.45)	(0.02)	(52.68)	(1.32)	(0.02)	(52.45)
2019×R	-0.80*	-0.00	32.52	-0.09	0.01	56.76
	(0.47)	(0.02)	(58.96)	(0.55)	(0.02)	(53.18)
2019×RV	-0.81*	-0.02*	88.44	-0.30	-0.00	149.52***
	(0.40)	(0.01)	(54.55)	(0.21)	(0.01)	(53.35)
Control mean	0.02	0.01	448.61	0.02	0.007	448.61
(Std. dev.)	(0.30)	(0.08)	(386.97)	(0.30)	(0.08)	(386.97)
N	3,434	3,456	3,273	4,123	4,144	4,002
R-squared	0.239	0.398	0.045	0.226	0.383	0.050
Village FE	YES	YES	YES	YES	YES	YES

Notes: V denotes Voucher group, R denotes the Recommendations group, and RV denotes the Recommendations and Voucher group. Robust standard errors in parentheses. Standard errors are clustered at the village level.
*** p<0.01, ** p<0.05, * p<0.1

Because 52% of the 2019 MMPs are new plots that were not tested as a part of the original RCT (Table 6), and given the lack of fertilizer application, it seems that this treatment impact may be driven by measurement error in productivity, which we explore in the following section.

5.3 Satellite-Derived Productivity

In the analysis that follows, we investigate measurement error in productivity as a possible factor driving the treatment impact on maize yields. We restrict the sample to plots that have both satellite-based and SR yields in order to compare results. Furthermore, because satellite-based yield measures in intercropped fields are a combined measure of productivity on all crops, and thus are less directly comparable to SR yields of a single crop (maize), we include an indicator variable that equals unity if a plot is pure-stand and 0 if it is intercropped. From the 920 original surveys, we are left with 659 and 409 observations for the 2014 MMP in 2016 and 2019, respectively, 699 observations for the 2016 MMP, and 689 observations for the 2019 MMP, for which we have both satellite and SR yields.

Table 9 displays the results of estimating equation (5) - panel A reports results for the 2014 MMP while panel B reports the 2016/19 MMP results. We report in the first six columns results without village FE, but we include these in the last six columns. This is not optimal from the broader goal of being able to use satellites on their own, but it helps explain why satellite-based estimates and SR estimates may give different results. As can be seen from both panels, when we restrict the satellite-based and SR estimates to the same sample, the R+V treatment does not have a significant impact using SR data in 2016 when we do not control for baseline imbalance using village FE (column 1 in Table 9), but it has a significant impact when we include village FE (column 7). In 2019, the R+V treatment leads to higher yields using SR data for the 2019 MMP only, with an amount of around 80 kg/acre, which represents around 18% increase from the control group's mean.

Moving to remote sensing measures, the satellite estimates indicate no treatment effect in any year on any of the MMPs, with two exceptions: we see in column 2 of Table 9 that the V treatment increased the 2014 MMP's productivity in 2016 by around 4% from the control group's mean. However, when we control for village-specific heterogeneity in column 8, this treatment impact disappears. The second instance in which we observe a significant treatment impact using satellite estimates is an increase of around 3.5% for the R+V group's yields obtained from the 2019 MMP (column 5). Including village FE, however, makes this impact vanish. Therefore, based on a more objective measure of productivity, we do not see a robust or substantial treatment impact.

We also use seemingly unrelated regressions (SUR) after applying an inverse hyperbolic sine (IHS) transformation for the outcomes in order to compare SR and satellite data. In columns 3 and 6 (Table 9), we report results of testing for SUR coefficient equality in 2016 and 2019 without village FE, while columns 9 and 12 display the same comparisons after including village FE. Although the coefficients vary in their signs and magnitudes, the only significant estimates are those of the 2019 MMP for the V group, with a difference of 0.46 (column 6) without controlling for village-specific heterogeneity and 0.43 (column 12) when controlling for it. Therefore, the DD V treatment impact using SR data observed in Table 8 on the 2019 MMP (which represents an increase of 22%) may be driven by an over-estimation of productivity, pointing to the possibility of measurement error in SR productivity.

Table 9: Treatment Effect on Self-Reported and Satellite-Derived Productivity

	(1) SR 2016	(2) Satellite 2016	(3) IHS SUR 2016	(4) SR 2019	(5) Satellite 2019	(6) IHS SUR 2019	(7) SR 2016	(8) Satellite 2016	(9) IHS SUR 2016	(10) SR 2019	(11) Satellite 2019	(12) IHS SUR 2019
Panel A: 2014 MMP												
V	15.68	0.09*	0.03	-9.84	-0.01	0.42	50.29	0.02	0.12	45.88	-0.04	0.37
	(53.23)	(0.04)	(0.25)	(44.19)	(0.05)	(0.27)	(45.64)	(0.03)	(0.27)	(37.22)	(0.05)	(0.29)
R	-39.32	0.03	-0.23	-5.31	0.01	0.29	-4.66	-0.02	-0.18	21.61	-0.01	0.20
	(40.43)	(0.05)	(0.26)	(49.07)	(0.07)	(0.29)	(29.02)	(0.05)	(0.29)	(56.53)	(0.07)	(0.30)
R+V	80.19	0.06	-0.14	9.88	0.01	0.07	118.38***	0.02	-0.01	53.29	-0.02	0.14
	(53.99)	(0.05)	(0.25)	(51.57)	(0.05)	(0.28)	(42.94)	(0.04)	(0.28)	(41.24)	(0.05)	(0.30)
Pure stand (=1)	-35.69	-0.08		33.83	-0.02		-75.88	-0.06		52.34	-0.02	
	(50.68)	(0.06)		(40.19)	(0.04)		(46.35)	(0.06)		(46.65)	(0.04)	
Control mean	481.02	2.18		452.35	2.45		481.02	2.18		452.35	2.45	
(Std. dev.)	(397.97)	(0.42)		(347.12)	(0.41)		(397.97)	(0.42)		(347.12)	(0.41)	
N	659	659	659	409	409	409	659	659	659	409	409	409
R-squared	0.009	0.009		0.002	0.001		0.017	0.004		0.008	0.002	
Village FE	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES
Panel B: 2016/19 MMP												
V	13.26	0.05	-0.05	34.58	0.04	0.46**	53.32	0.01	0.11	47.80	-0.01	0.43*
	(48.07)	(0.04)	(0.24)	(42.15)	(0.05)	(0.22)	(44.71)	(0.03)	(0.26)	(32.76)	(0.05)	(0.23)
R	-51.00	0.00	-0.29	25.84	0.03	0.29	-16.23	-0.04	-0.22	35.46	-0.02	0.24
	(37.86)	(0.05)	(0.25)	(37.72)	(0.04)	(0.22)	(30.71)	(0.05)	(0.27)	(39.65)	(0.04)	(0.23)
R+V	72.57	0.02	-0.12	81.47*	0.08**	0.33	108.81***	-0.00	0.01	82.58**	0.02	0.30
	(49.73)	(0.05)	(0.24)	(40.86)	(0.04)	(0.22)	(37.75)	(0.05)	(0.26)	(32.84)	(0.03)	(0.23)
Pure stand (=1)	-19.91	-0.07		73.31**	0.02		-67.34	-0.05		73.67**	0.03	
	(51.78)	(0.05)		(29.42)	(0.04)		(48.92)	(0.06)		(32.47)	(0.03)	
Control mean	474.33	2.20		452.89	2.40		474.33	2.20		452.89	2.40	
(Std. dev.)	(394.18)	(0.43)		(335.63)	(0.41)		(394.18)	(0.43)		(335.63)	(0.41)	
N	699	699	699	689	689	689	699	699	699	689	689	689
R-squared	0.009	0.004		0.018	0.007		0.016	0.003		0.017	0.003	
Village FE	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES

Notes: V denotes Voucher group, R denotes the Recommendations group, and R+V denotes the Recommendations and Voucher group. Control mean refers to the mean of the control group at baseline, unless in the case of satellite estimates, for which we do not have baseline data and hence report endline averages. Robust standard errors in parentheses. Standard errors are clustered at the village level.
*** p<0.01, ** p<0.05, * p<0.1

5.4 Correlation between Self-Reported and GPS-Estimated Plot Size

We start by presenting in Table 10 difference in means between the two measures of plot size – SR and GPS estimates – using paired t-tests. In panel A we report results for the 2014 MMP, while panel B displays the 2016/19 MMPs' results. All differences between SR and GPS areas are negative and distinguishable from zero at the 1% level, indicating that farmers, on average, under-report plot size, resulting in inflated productivity. Measurement error in Table 10 is smallest for the R+V farmers and largest for the R farmers, both in absolute value (column 5) and relative to the GPS-based true plot size (column 7). Compared to the control group, there does not seem to be a substantial difference in measurement error among the three treatments.

Table 10: Measurement Error in Plot Size by Treatment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Self-Reported		GPS-estimated		Measurement error		Relative	N
	Mean	SD	Mean	SD	Mean	SD	bias	
Panel A: 2014 MMP								
C	1.72	1.42	2.03	1.86	-0.314***	1.45	-15.27%	845
V	1.7	1.38	2.07	1.92	-0.368***	1.34	-17.87%	380
R	1.76	1.9	2.16	2.27	-0.409***	1.92	-18.98%	324
R+V	1.78	1.53	2.05	1.44	-0.271***	1.43	-13.17%	360
Total	1.73	1.52	2.07	1.88	-0.333***	1.52	-15.94%	1909
Panel B: 2016/19 MMP								
C	1.71	1.42	1.99	1.83	-0.279***	1.36	-14.07%	1003
V	1.78	1.69	2.13	2.27	-0.349***	1.43	-16.43%	434
R	1.72	1.82	2.08	2.17	-0.354***	1.82	-16.83%	378
R+V	1.76	1.51	2.01	1.45	-0.252***	1.35	-12.44%	420
Total	1.74	1.56	2.04	1.92	-0.300***	1.46	-14.71%	2235

Notes: C denotes the Control group, V denotes Voucher group, R denotes the Recommendations group, and R+V denotes the Recommendations and Voucher group. Robust standard errors in parentheses. Standard errors are clustered at the village level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Next, we estimate the correlation between SR and GPS-based areas using equation (6). The 2014 MMP's results are reported in panel A in Table 11, while panel B displays the 2016/19 MMPs' results. Recall that if the slope is around unity and the regression constant is around zero, then SR and GPS estimates are very similar. When we pool all treatments together, we see that the slope and the constant have a similar value (column 1 in Table 11). However, this aggregation masks differences between treatments. When looking at the correlation between the GPS estimates and SR plot size for the R+V farmers, we notice that this correlation is lowest since the slope is smallest among all treatments with a magnitude of 0.5, and the constant term is largest with a magnitude of 1.2 for the 2014 MMP. In panel B, we again notice that the R+V farmers still have the largest constant term and the smallest slope.

5.5 Productivity Using Self-Reported vs. GPS-Based Plot Size

We examine if the measurement error in SR areas has a role in the treatment impact on the R+V farmers by comparing productivity (kg/acre) using SR production in kilograms per acre planted, where acres are reported either by the respondent or derived via GPS estimates. Since we have baseline production data, we use the same preferred DD

Table 11: Correlation Between GPS-Estimated and Self-Reported Plot Size

	(1) All	(2) C	(3) V	(4) R	(5) R+V
Panel A: 2014 MMP					
SR plot size	0.76***	0.84***	1.00***	0.70***	0.50***
	(0.06)	(0.06)	(0.07)	(0.17)	(0.13)
Constant	0.74***	0.60***	0.37**	0.93***	1.16***
	(0.13)	(0.12)	(0.16)	(0.27)	(0.24)
N	1,909	845	380	324	360
R-squared	0.385	0.408	0.515	0.346	0.286
Panel B: 2016/19 MMP					
SR plot size	0.82***	0.87***	1.05***	0.71***	0.56***
	(0.05)	(0.06)	(0.06)	(0.16)	(0.13)
Constant	0.61***	0.50***	0.26**	0.85***	1.03***
	(0.11)	(0.11)	(0.11)	(0.25)	(0.23)
N	2,235	1,003	434	378	420
R-squared	0.446	0.459	0.607	0.355	0.339

Notes: C denotes the Control group, V denotes Voucher group, R denotes the Recommendations group, and R+V denotes the Recommendations and Voucher group. Robust standard errors in parentheses. Standard errors are clustered at the village level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

estimator employed in equation (2), but we restrict the sample to plots that have both GPS-based and SR areas. We also restrict the analysis to the years to 2014, 2016, and 2019 because these are the years for which we have GPS areas.[6] The results are reported in Table 12.

The first four columns report the 2014 MMP's results while the last four columns display the 2016/19 MMPs' results. In columns 3-4 and 7-8, we restrict the sample to plots that have GPS size of more than 0.5 acres since including small plots can reduce the coefficient of determination (Burke & Lobell, 2017). Interestingly, Table 12 shows that when GPS areas are used, the treatment impact on yields of farmers in the R+V group in 2016 vanishes (columns 2, 4, 6, 8). The treatment impact on yields obtained from the 2019 MMP for farmers in the R+V group, however, is persistent (only at the 10% level) when we use GPS estimates (columns 6 and 8 in Table 12), which suggests that the largest difference between SR and GPS areas takes place in 2016 relative to 2014. Measurement error in areas might help explain the treatment impact, if we assume GPS areas are more accurate than SR ones.

The results reported in Table 12 also indicate that farmers in the V group have higher productivity in 2019 derived from the 2014 MMP when its plot size is larger than 0.5 acres (columns 3 and 4), and also from the 2019 MMP, both for the entire sample and when restricted to plots that have a size of more than 0.5 acres (columns 5-8). The satellite results in Table 9 showed that farmers in the V group over-report their 2019 MMP's productivity. In this analysis, we cannot quantify measurement error in output, but if we assume that farmers in the V group over-report output especially that they were shown to over-estimate productivity in Table 9, then over-reporting output may also lead to inflated productivity despite correcting for measurement error in plot size, explaining the significant V treatment

[6] Since we did not collect baseline data on SR output and area of the 2014 MMP as a part of the original study, we relied on productivity of the 2014 MMP using output and areas that were measured by agronomists. In this analysis, we follow the same procedure and measure productivity at baseline as estimated by agronomists.

impact. Indeed, using an objective measure of plot size has reduced the V treatment impact from 117 kg/acre to 99 kg/acre (columns 5 and 6 in Table 12), but using this objective plot size does not rule out the possibility of over-reporting output, which we discuss in more detail in section 6.

5.6 Robustness of Results

l We test the robustness of our results by first running the same analysis where we pool the R and control groups together, and the V and the R+V groups together. The reason behind this pooling is that all farmers were given the recommendations after 2016; thus, the voucher distinguishes the groups over 2017-2019. First, the DD results indicate that farmers in the pooled V and R+V group have higher SR yields in all years and from both of their 2014 MMP and 2016-2019 MMPs (Table 22 in Appendix A.3), and they apply more fertilizer in 2017 but not in 2018 and 2019, suggesting that the original DD results are robust to pooling.

Second, moving to the OLS regressions using SR and satellite-derived yields and focusing on our preferred specifications that include village FE, the treatment impact using satellite estimates stays insignificant when we pool the V and R+V groups. As for the treatment impact on the 2014 MMP's SR yields, it declines from 118 kg/acre to 87 kg/acre in 2016 (column 5 in Table 23) but changes from being insignificant in 2019 to significant at the 10% level with a magnitude of 39 kg/acre (column 7 in Table 23). Similarly, the 2016 MMP's SR productivity declines from 109 kg/acre to 89 kg/acre (column 5 in Table 23) and that of the 2019 MMP from 83 kg/acre to 48 kg/acre Therefore, it seems that pooling reduces the SR treatment impact.

Third, the pooled DD results for comparing SR and GPS-based productivity indicate that the treatment impact on SR productivity is smaller in magnitude (by around 20-25 kg/acre) when we pool the V and R+V groups for all plots and years, but still significant (Table 24). As for GPS-based productivity, it stays insignificant for the 2014 MMP in 2016, and significant at the 10% level in 2019 with similar magnitudes. The 2016 MMPs' GPS-based productivity, however, becomes significant when the sample is restricted to plots of size more than 0.5 acres with a magnitude of 81 kg/acre, and that of the 2019 MMP stays significant but is reduced by around 15 kg/acre. Overall, it seems that pooling reduces some of the coefficients, but makes some previously insignificant impacts more precise.

As a second robustness check, in Appendix A.4, we use ANCOVA shown in equation (3), which may be thought of as a lower bound of the treatment impact. As Table 25 shows, the impact in 2016 is now reduced from 167 kg/acre to 123 kg/acre for the 2014 MMP and from 152 kg/acre to 105 kg/acre for the 2016 MMP. However, the remaining results over 2017-2019 become insignificant for all plots, except for the 2019 MMP's SR R+V productivity that declines from 150 kg/acre to 85 kg/acre. We check the robustness of the GPS-based productivity by employing ANCOVA again (Table 26). In 2016, only SR R+V productivity is significant but is smaller in magnitude by 40-50 kg/acre, while GPS-based productivity stays insignificant. Moving to 2019, all of the V treatment impacts become insignificant, while the GPS-based R+V treatment effect is reduced from 138 kg/acre to 95 kg/acre for the entire sample, and becomes insignificant when only plots of size larger than 0.5 acres are included. Overall, by using ANCOVA, the SR treatment impact in 2016 (on all plots) and 2019 (2019 MMP) stays significant but becomes smaller, whereas GPS-based productivity is reduced in magnitude by around 30%. Table 27 provides a summary of all the results.

Table 12: Treatment Effect on Productivity Using Self-Reported and GPS-Estimated Plot Size

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2014 MMP		>0.5 acres		2016/19 MMP		>0.5 acres	
	SR	GPS	SR	GPS	SR	GPS	SR	GPS
2016	-141.27***	-120.97***	-136.92***	-130.28***	-144.29***	-119.48***	-140.68***	-137.55***
	(38.02)	(38.43)	(39.56)	(35.22)	(35.82)	(37.67)	(37.43)	(33.84)
2019	-127.16***	-106.96***	-135.25***	-112.14***	-142.20***	-101.37***	-151.31***	-108.37***
	(35.75)	(35.47)	(29.78)	(31.51)	(29.76)	(30.87)	(27.38)	(28.99)
V	-28.67	-29.29	-21.45	-41.26	-44.87	-37.54	-39.74	-50.63*
	(29.26)	(30.50)	(28.71)	(30.39)	(29.93)	(28.53)	(29.65)	(27.98)
R	35.99	-7.16	44.26	-9.95	19.73	-16.80	25.74	-20.61
	(49.22)	(47.36)	(50.71)	(47.92)	(48.98)	(45.51)	(50.37)	(45.96)
R+V	-43.71	-19.83	-47.55	-31.39	-59.73*	-27.65	-64.55**	-40.64
	(31.47)	(52.35)	(32.28)	(50.30)	(31.38)	(50.35)	(31.81)	(47.60)
2016×V	92.65	52.76	82.96	69.40	94.22	50.97	86.13	77.32
	(61.75)	(52.75)	(63.80)	(51.97)	(58.56)	(49.48)	(60.73)	(48.04)
2016×R	-32.56	-22.39	-47.13	-29.07	-39.15	-24.13	-51.38	-21.20
	(56.92)	(54.91)	(57.90)	(50.93)	(52.18)	(52.92)	(53.65)	(48.17)
2016×RV	158.18**	56.44	163.39**	64.78	155.02***	55.54	160.09***	73.21
	(62.09)	(67.43)	(62.80)	(59.63)	(57.16)	(66.55)	(57.90)	(58.76)
2019×V	81.44	85.65	81.89*	95.43*	117.27**	98.52**	119.45**	112.16**
	(50.13)	(52.68)	(44.53)	(50.10)	(47.58)	(45.23)	(47.06)	(44.71)
2019×R	7.16	8.66	1.32	11.21	42.96	63.84	42.81	60.42
	(61.89)	(53.31)	(59.25)	(53.02)	(57.19)	(62.24)	(60.02)	(61.40)
2019×RV	90.05	70.88	100.50*	60.47	169.62***	137.52*	181.07***	116.30*
	(56.98)	(69.43)	(53.90)	(58.73)	(51.01)	(75.73)	(48.87)	(57.92)
Control mean	455.23	417.92	456.98	402.67	455.23	417.92	456.98	402.67
(Std. dev.)	(385.24)	(414.67)	(390.47)	(394.80)	(385.24)	(414.67)	(390.47)	(394.80)
N	1,817	1,817	1,700	1,700	2,137	2,137	1,981	1,981
R-squared	0.026	0.019	0.028	0.025	0.027	0.018	0.030	0.024
Village FE	YES	YES	YES	YES	YES	YES	YES	YES

Notes: V denotes Voucher group, R denotes the Recommendations group, and RV denotes the Recommendations and Voucher group. Control mean refers to the mean of the control group at baseline. Robust standard errors in parentheses. Standard errors are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

6 Discussion

Given the distinct data sources, the R+V treatment seems to have had a varying impact on productivity, with self-reports indicating an increase of 23%-37% in 2016, no impact on the 2014 MMP in 2019, and a 19%-33% increase on the 2019 MMP. On the other hand, satellite images reveal a minor or no treatment impact both in 2016 and 2019, while GPS-based measures indicate no treatment effect in 2016 on all plots, but an effect that ranges between 23%-33% on the 2019 MMP. This latter case represents a disagreement between satellite-derived and GPS-based productivity because the 2019 MMP's productivity is positive and significant using GPS-based estimates. A possible explanation for this apparent disagreement is that farmers over-report output, as has been shown to be the case in Gourlay, Kilic, and Lobell (2019), when SR maize output was compared to crop cutting–based estimates, the gold standard measure of output. To validate this argument, we asked farmers in 2019 to recall the area and output of their 2014 MMP in 2016 and their 2016 MMP, allowing us to investigate recall bias by comparing productivity in 2016 using SR actual data collected in 2016 from Harou et al. (2020) and 2016 SR data recalled in 2019. Figures 8 and 9 present results for the 2014 and 2016 MMP, respectively.

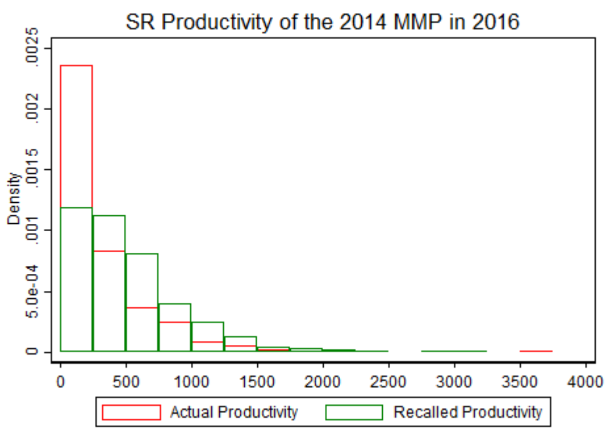

Fig 8: SR Productivity of the 2014 MMP (in kg/acre)

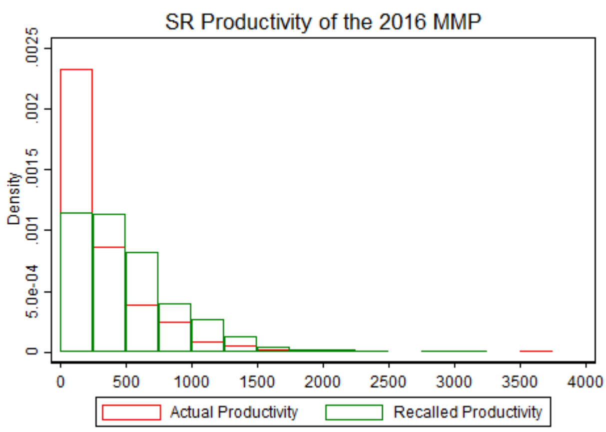

Fig 9: SR Productivity of the 2016 MMP (in kg/acre)

As can be seen from both figures, respondents tend to over-report productivity (in kg/acre) on their plots when asked in 2019 to recall the 2016 size and production. Although the data indicate that farmers over-report their plot size by around half an acre on average when they recall the area three years later, they also over-report output by around 620-640 kg, which is so high that it drives a wedge between actual and recalled productivity of more than 200 kg/acre (around a 70% increase) on average. Given these vast differences when investigating one form of measurement error that stems from recall bias, it is possible that farmers tend to over-estimate their production, leading to the R+V treatment impact on the 2019 MMP, even when we control for measurement error in plot size. Although this analysis does not explain fully the treatment impact observed in 2019 (even after controlling for measurement error in plot size) because the 2019 data do not suffer from recall bias, assuming other forms of measurement error, such as misperceptions (Abay et al., 2020), follow a similar pattern, then this recall bias may represent an evidence of over-reporting output in general, a finding confirmed in Gourlay et al. (2019).

Even if we assume that the R+V treatment has had an impact on productivity and profits as reported by farmers (Harou et al., 2020), it remains that they do not apply fertilizer after the original intervention was concluded in 2016, which reconciles all of the different productivity results as farmers seem not to be convinced of the profitability of fertilizer, given their limited resources. To examine this argument, we explore three different reasons that may explain why farmers choose not to invest in fertilizer. First, are farmers unable to recall the recommendations that were given to them? Second, do farmers (incorrectly) believe that their soils are fertile and hence do not require fertilizer? Third, are farmers not willing to pay for fertilizer?

6.1 Information Retention

To test the effect of the V, R, and R+V treatments on farmers' information retention, we construct standardized retention indices, with higher values indicating better retention abilities. The indices are based on farmers' ability to recall the recommendations in Table 13, which shows the recommended fertilizers based on the soil nutrient limitations for the 1007 farmers whose soils were tested. The question that we asked farmers in 2016 was "Can you please list the types of fertilizer that were recommended to you [source: SUA SoilDOC project]." and that in 2019 was "Can you please list the types and quantities of ALL fertilizers (basal and top dressing) PER ACRE that were recommended to you for your 2014 main maize plot?"

Table 13: Fertilizer Recommendations

Nutrient limitations	Treatment	#	%
N	25 kg Urea + 2x5 kg Urea	36	3.57
NS	25 kg SA + 25 kg Urea	639	63.46
NP	25 kg DAP + 25 kg Urea	7	0.7
NPS	40 kg Minjingu Mazao + 25 kg Urea	215	21.35
NK	25 kg MOP + 25kg Urea + 2x5 kg Urea	2	0.2
NKS	25 kg SOP + 25kg Urea + 2x5 kg Urea	56	5.56
NPK	25 kg DAP + 25kg MOP + 25 kg Urea	2	0.2
NPKS	50 kg Yara Mila Tobacco + 25 kg Urea	50	4.97
Total		1007	100

Notes: The above recommendations are for a land area of half an acre planted with maize. N is nitrogen, S is sulfur, P is phosphorus, and K is potassium. SA stands for ammonium sulfate, DAP for diammonium phosphate, MOP for muriate of potash (potassium chloride), and SOP for sulfate of potash (potassium sulphate).

Since we do not have data on recalled fertilizer quantities in 2016, the 2016 index is constructed by first assigning one point for every fertilizer type recalled correctly. Then, we subtract the mean and divide by the standard deviation (SD) to normalize it. As for 2019, we construct three retention indices by assigning scores and then standardize them. The first measure assigns one point for every fertilizer type recommended that was correctly recalled and disregards the quantity in order to mimic the 2016 index; the second measure assigns one point for every fertilizer type recommended and two points for every fertilizer quantity recommended that were correctly recalled, assuming that it is harder to recall quantities; in the third index, each farmers is given one point for each type and one point for each quantity recalled correctly, thus relaxing the assumption that it is harder to recall quantities. For farmers who claim

Table 14: Raw Retention Scores

	The 2016 scores	First 2019 scores	Second 2019 scores	Third 2019 scores
0	737	546	546	546
1	126	280	127	127
2	121	94	30	183
3	-	-	153	16
4	-	-	16	48
6	-	-	48	-
N	984	920	920	920

not to have received recommendations or who do not recall the recommendations, we assign them a score of zero.[7] In Table 14, we present the raw retention scores before normalization.

After creating the index, we study retention by estimating equation (5) in 2016 and again in 2019 using OLS since retention is a post-intervention measure. The results are shown in Table 15. The first column of Table 15 shows the results for the 2016 index, while the remaining columns show the 2019 indices. Compared to the control group, farmers in the R and R+V groups have one additional SD and 1.5 SD in 2016, respectively, and this effect is significant at the 1% level. These results are to be expected because farmers in the control and V groups received their plot-specific recommendations only after the 2016 endline data were collected. Indeed, in 2016, there is no substantial effect on retention for farmers in the V group (0.1 SD) compared to farmers in the control group. When measured in 2019, retention of farmers in the R group becomes indistinguishable from zero, while the increase in scores of farmers in the R+V group that was observed in 2016 drops to around 0.4 SD in in all of the three indices. In 2019, there is still no difference in retention between farmers in the V group and the control group. This finding is unexpected because farmers in the V group received their recommendations after 2016, more recently than the other farmers, but highlights the importance of receiving information and using it simultaneously so that this information is not disregarded.

Having the opportunity to practice a technology, i.e., applying fertilizers, appears to help farmers retain the information provided to them. Indeed, since farmers in the R group were not given vouchers to use the information, and since farmers in the V group were given vouchers before being given the recommendations, they were also unable to act on the information. Since the R+V farmers recall the recommendations in 2019 but to a lesser extent compared to 2016, we can rule out the possibility that farmers are unable to recall the recommendations as an explanation for the lack of adoption, pointing again to the idea that farmers are not convinced of the profitability of fertilizer, given the risks associated with adopting.

[7]Since 3.6% of farmers were recommended to apply urea only (N-limited group in Table 13), we double their scores to make them comparable to the majority of farmers who were recommended two fertilizers. Only two farmers were recommended to apply three fertilizers and neither of them recalled the recommendations in 2019, so their scores are zero. However, one farmer of them was able to recall the three fertilizers in 2016, and so we change the score from three to two points to be consistent with the majority of farmers who were recommended to apply two fertilizers.

Table 15: Treatment Effect on Information Retention

	The 2016 retention index	First 2019 retention index	Second 2019 retention index	Third 2019 retention index
V	0.10**	0.09	0.09	0.09
	(0.05)	(0.09)	(0.08)	(0.08)
R	1.01***	0.05	0.02	0.03
	(0.13)	(0.08)	(0.08)	(0.08)
R+V	1.50***	0.45***	0.39***	0.42***
	(0.09)	(0.10)	(0.11)	(0.11)
Control mean	-0.51	-0.12	-0.09	-0.10
(Std. dev.)	(0.26)	(0.94)	(0.97)	(0.96)
N	984	920	920	920
R-squared	0.337	0.025	0.019	0.022
Village FE	YES	YES	YES	YES

Notes: V denotes Voucher group, R denotes the Recommendations group, and R+V denotes the Recommendations and Voucher group. Robust standard errors in parentheses. Standard errors are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

6.2 Soil Fertility Beliefs

6.2.1 A Bayesian model of updating beliefs

In their work, Murphy, Roobroeck, Lee, and Thies (2020) incorporate both an empirical and a theoretical approach to farmers updating their willingness to pay (WTP) for fertilizer upon receiving soil information. The authors motivate their study with a Bayesian theoretical model in which farmers update their prior beliefs about input profitability. They conduct a randomized control experiment to elicit Kenyan farmers' WTP for DiAmmonium Phosphate (DAP) fertilizer. Their results indicate that WTP for DAP increased when DAP was recommended, highlighting the importance of soil testing to increase WTP for fertilizer with the potential for increasing adoption. Compared to their study, we assess the longer-term effects of providing farmers with soil information on soil perceptions and not on an auction-based WTP.

We first present a simple model of farmers updating their soil fertility perceptions following Bayesian updating. In this study, farmers are assigned randomly to a vouchers-only treatment, a recommendations-only treatment, both treatments, or none (control). The information about true deficiencies can cause farmers to either not change their prior beliefs, or to revise their priors upwards/downwards. Conditional on acting on that information provided, farmers might believe that they are improving their soils because they are addressing the deficiency identified.

We proceed to describing the model. Let SSB_{it}, the subjective soil belief of farmer i at time t that varies based on the soil nutrient content, be unknown to farmers precisely, but its estimates follow a normal distribution:

$$SSB_{it} \sim N\left(\overline{SSB}_{it}\,, \nu_{it}^2\right) \tag{7}$$

Define the accuracy of a farmer's belief, ψ_{it}, by the inverse of its variance to reflect that a farmer's belief about the soil quality is more accurate if it is less variable:

$$\psi_{it} = \frac{1}{\nu_{it}^2} > 0 \tag{8}$$

A farmer receives an unbiased information signal R from our recommendation treatment about the true soil quality, which includes information about the soil's nutrient deficiencies and the suitable fertilizer to apply. A farmer might also receive a signal V from the voucher treatment that permits purchasing fertilizer, increasing the true soil quality through the nutrients coming from the applied fertilizer. Assuming the signal K (R or V) to be normally distributed, we have:

$$K \sim N\left(\xi_{it}^{K},\ \sigma_{it}^{2(K)}\right), K = R, V \tag{9}$$

If a farmer updates his/her soil beliefs following Bayesian updating, posterior mean SSB is given by the weighted average of the prior beliefs and the signal received:

$$\overline{SSB}_{it+1}^{K} = \frac{\nu_{it}^{2}}{\nu_{it}^{2} + \sigma_{it}^{2(K)}}\xi_{it}^{K} + \frac{\sigma_{it}^{2(K)}}{\nu_{it}^{2} + \sigma_{it}^{2(K)}}\overline{SSB}_{it} = \frac{\xi_{it}^{K} + \psi_{it}\sigma_{it}^{2(K)}\overline{SSB}_{it}}{1 + \psi_{it}\sigma_{it}^{2(K)}}, K = R, V \tag{10}$$

Notice that in equation (10), posterior and prior soil beliefs are unchanged if the signal received, whether bad or good, is equal to the prior beliefs:

$$\xi_{it}^{K} = \overline{SSB}_{it} \Longrightarrow \overline{SSB}_{it+1}^{K} = \overline{SSB}_{it},\ K = R, V \tag{11}$$

Based on this model, there are three propositions that we can test empirically:

Proposition 1:

Upon receiving the R treatment, and based on equation (10), we have three possible cases:

$$\xi_{it}^{R} < \overline{SSB}_{it} \Longrightarrow \overline{SSB}_{it+1}^{R} < \overline{SSB}_{it}, \tag{12}$$

$$\xi_{it}^{R} > \overline{SSB}_{it} \Longrightarrow \overline{SSB}_{it+1}^{R} > \overline{SSB}_{it}, \tag{13}$$

$$\xi_{it}^{R} = \overline{SSB}_{it} \Longrightarrow \overline{SSB}_{it+1}^{R} = \overline{SSB}_{it}, \tag{14}$$

Equation (12) means that in the case of an R treatment, a farmer would update his/her prior beliefs downwards if the information about the soil quality received is lower than his/her prior beliefs i.e. what he/she thinks of his/her soil fertility initially. However, a farmer would not update his/her belief if the bad signal is equal to his/her prior belief about the soil quality, as seen in equation (14)

Proposition 2:

From (10), when a farmer receives a voucher treatment, we have:

$$\xi_{it}^{V} > \overline{SSB}_{it} \Longrightarrow \overline{SSB}_{it+1}^{V} > \overline{SSB}_{it}, \tag{15}$$

Equation (15) means that when a farmer receives a V treatment, he/she will update his/her prior belief upwards if the V treatment has a larger impact on SSB than the prior belief. This scenario happens because vouchers allowed purchasing fertilizers that replenished the soil's nutrient content, meaning that farmers were able to act and apply fertilizer.

Proposition 3:

When V and R signals are given simultaneously, there are three possible cases:

$$\xi_{it}^{V} < \xi_{it}^{R} \implies \overline{SSB}_{it+1}^{RV} < \overline{SSB}_{it} \tag{16}$$

$$\xi_{it}^{V} > \xi_{it}^{R} \implies \overline{SSB}_{it+1}^{RV} > \overline{SSB}_{it} \tag{17}$$

$$\xi_{it}^{V} = \xi_{it}^{R} \implies \overline{SSB}_{it+1}^{RV} = \overline{SSB}_{it} \tag{18}$$

Therefore, farmers will update their prior SSB downwards (upwards) if they believe the R (V) treatment to be stronger than the V (R) treatment, and they will not update their beliefs if the two treatments have comparable effect on prior SSB. These three propositions can be tested empirically using data on prior and posterior beliefs upon receiving V and/or R treatments compared to a control group of farmers who do not receive any treatment by 2016.

6.2.2 Testing the model

To test the model, we employ our preferred DD estimation in equation (2) because we have baseline data. Recall that subjective soil beliefs (SSB) is elicited by asking farmers directly about their soil perceptions without being given any guidance, and is defined as an ordinal variable that takes a value of zero to indicate poor soil perceptions, one if a farmer has fair perceptions, and two indicates a good belief about the soil fertility. We restrict the sample to the years 2014, 2016 and 2019 to use the SSB data when they were collected (baseline in 2014, endline in 2016, and extension in 2019) to avoid recall bias. We also restrict the analysis to the 2014 MMP since this is the plot that was tested, and also because we asked farmers about their SSB in 2019 regardless of whether or not they cultivated the 2014 MMP.[8]

We are also interested in learning whether the effect of the R, V, and R+V treatments on SSB depends the underlying soil quality. In other words, we want to know whether farmers update their SSB conditional on the content of the information they receive. The left-hand side of equation (2) becomes $SSB|TSQ_i$ where TSQ stands for true soil quality, which is proxied through a soil index we create from seven soil fertility indicators (pH, electrical conductivity [EC], active carbon [AC], nitrogen [N], sulfur [S], potassium [K] and phosphorus [P]) using principal component analysis (PCA). The weights assigned to the indicators based on PCA are 0.23 for pH, 0.27 for EC and N, 0.19 for AC, and 0.31 for P, S, and K. The index takes a value between zero and one, with higher values indicating better soil

[8]The correlation between SSB of the 2014 and 2016/2019 MMP is 93% and the results that we will show are almost the same when we use the 2016/2019 MMP instead of the 2014 MMP.

Table 16: Descriptive Statistics of the TSQ by Limitation

Limitations	Mean	SD	Min	Max	#
N	0.78	0.10	0.58	1	36
NS	0.61	0.10	0.28	0.92	637
NP	0.50	0.13	0.35	0.73	7
NK	0.49	0.13	0.40	0.59	2
NPK	0.46	0.22	0.31	0.62	2
NPS	0.4	0.12	0.12	0.73	215
NKS	0.38	0.09	0.2	0.55	54
NPKS	0.20	0.11	0	0.45	50
Total	0.54	0.17	0	1	1003

Notes: Limitations refer to soil deficiencies classified by agronomists and soil scientists as a part of the 2014 experiment. N is nitrogen, S is sulfur, P is phosphorus, and K is potassium.

fertility.[9] Table 16 provides summary statistics of the TSQ index by nutrient deficiencies. In order to classify soils, we consider a soil score to be poor if it is less than 0.33, fair if it is higher than or equal to 0.33 and less than 0.66, and good if it is higher than or equal to 0.66. Based on this classification, 12.16 percent of farmers have "poor" TSQ, 61.81 percent have "fair" TSQ, and 26.02 percent have "good" TSQ.

The index is similar to that of Lobell et al. (2019) and Gourlay et al. (2019). It controls for pH, EC, AC, as well as other nutrients (S, K, N, and P). These variables are necessary in this study's context since most of the soils are deficient in N and S, and as there is a large variability in P and K content. Compared to Mukherjee and Lal (2014), this index takes into account nutrient supply capacity but due to data limitations does not incorporate two components of soil quality that they consider – root development and water storage capacities. More details on the construction of the soil index are provided in Appendix A.5. The results are reported in Table 17.

Column 1 in Table 17 shows the unconditional results. The DD coefficient is statistically different from zero only for farmers in the V group in 2016, with a magnitude of 0.17 when SSB is not conditioned on TSQ. Also, the DD coefficient for the R group in 2016 is negative (but insignificant) with a magnitude of 0.02 and increases to become 0.22 when SSB is conditioned on poor TSQ. Following the notation of the theoretical model, the upward update by the V farmers requires the mean of the vouchers signal received to be greater than prior beliefs:

$$\xi^V_{i2014} > \overline{SSB}_{i2014} \tag{19}$$

However, the absence of updating by farmers in the R group as indicated by the marginal and insignificant DD coefficient in 2016 (with an unconditional magnitude of -0.02) necessitates that:

$$\xi^R_{i2014} = \overline{SSB}_{i2014} \tag{20}$$

Hence, according to equations (19) and (20) and following the theoretical motivation, when both V and R signals are

[9]The average continuous soil score is 0.558 and standard deviation is 0.173.

Table 17: Treatment Effect on Subjective Soil Beliefs

	(1) SSB=0 poor; 1 fair; 2 good	(2) SSB|Poor TSQ	(3) SSB|Fair TSQ	(4) SSB|Good TSQ
2016	-0.00	-0.03	0.01	0.01
	(0.04)	(0.14)	(0.06)	(0.08)
2019	-0.15***	-0.15	-0.15**	-0.06
	(0.05)	(0.15)	(0.07)	(0.06)
V	-0.01	-0.13	-0.01	0.15**
	(0.05)	(0.19)	(0.05)	(0.06)
R	0.05	0.11	-0.01	0.13
	(0.06)	(0.15)	(0.07)	(0.12)
R+V	-0.01	-0.01	-0.02	0.08
	(0.04)	(0.18)	(0.06)	(0.09)
2016×V	0.17**	0.21	0.11	0.16
	(0.07)	(0.17)	(0.09)	(0.13)
2016×R	-0.02	-0.22	0.00	0.00
	(0.08)	(0.18)	(0.12)	(0.14)
2016×RV	0.09	0.19	0.03	0.10
	(0.08)	(0.20)	(0.11)	(0.13)
2019×V	0.07	-0.01	0.12	-0.09
	(0.09)	(0.21)	(0.10)	(0.19)
2019×R	0.02	-0.04	0.06	-0.12
	(0.10)	(0.21)	(0.13)	(0.16)
2019×RV	0.03	0.09	0.06	-0.15
	(0.06)	(0.20)	(0.11)	(0.11)
Control mean	1.62	1.48	1.61	1.65
(Std. dev.)	(0.59)	(0.72)	(0.61)	(0.54)
N	2,833	329	1,669	714
R-squared	0.017	0.029	0.012	0.031
Village FE	YES	YES	YES	YES

Notes: V denotes Voucher group, R denotes the Recommendations group, and RV denotes the Recommendations and Voucher group. Control mean refers to the mean of the control group at baseline. TSQ is the true soil quality index, based on which soils are classified as of poor, fair, or good fertility. Robust standard errors in parentheses. Standard errors are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

combined as was the case of farmers in the R+V group, it follows that:

$$\xi^{V}_{i2014} > \xi^{R}_{i2014} \implies \overline{SSB}^{RV}_{i2016} > \overline{SSB}_{i2014} \tag{21}$$

Equation (21) means that farmers in the R+V group should update their beliefs upwards as a result of receiving both recommendations and vouchers since the V farmers updated their prior SSB upwards, but the R farmers did not. When these implications are taken to the data, we observe in column 1 of Table 17 that equation (21) does not hold since γ_{RV2016} is insignificant. Therefore, assuming farmers update their prior beliefs following Bayesian updating, our results indicate that farmers do not update their prior beliefs. Furthermore, farmers seem not to update their beliefs based on the underlying true soil quality as all coefficients in columns 2-4 in Table 17 are insignificant, suggesting that farmers may be unaware of their soils' inadequate nutrient content.

Farmers in the R+V group observe higher SR yields from their 2014 MMP in 2016. Yet, their beliefs about the soil fertility are unchanged. On the other hand, farmers in the V group do not obtain additional SR yields in 2016, but they update their beliefs upwards. Studies by Marenya et al. (2008) and Berazneva et al. (2018) find yields from survey data to correlate positively with soil perceptions. However, using SR data, we do not find an evidence of this correlation, which is an important experimental contribution, highlighting that many factors may confound the relationship between yields and perceptions. Thus, it seems that farmers think that their soils are fertile and hence do not need fertilizer, especially that more than 65% of farmers in our sample ranked their soils' quality as "good" or "very good". Another possibility is that farmers question fertilizer's profitability, contributing to explaining farmers' lack of fertilizer application despite witnessing short-term gains on the level of SR productivity and profitability, as shown in Harou et al. (2020). In a study conducted in the Morogoro Region, the same area in our study, Michelson, Fairbairn, Ellison, Maertens, and Manyong (2021) find that farmers' incorrect beliefs about fertilizer quality affect their demand for fertilizer, highlighting the importance of farmers' erroneous beliefs in explaining fertilizer underuse.

6.3 Willingness to Pay for Recommendations

The third factor that we explore and that may play a role in explaining farmers' reluctance to invest in fertilizer is that they are not willing to pay for it. Two studies done in Tanzania – Michelson et al. (2021) and Shee, Azzarri, and Haile (2020) – find that farmers are not willing to pay for fertilizer at the prevailing market prices. Although we did not elicit WTP for fertilizer, we elicited WTP for tailored plot-specific soil recommendations using contingent valuation. If WTP for recommendations is unchanged, then this supports the argument that farmers are not convinced of the profitability of fertilizer because these recommendations are the factor that distinguishes the V and R+V groups, with the latter having higher SR yields. To test for this possibility, we estimate equation (5) in 2016 using OLS since WTP for soil information is a post-intervention measure. The results are reported in Table 18.

Table 18: Treatment Effect on WTP for Recommendations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	WTP indicator (=1 if willing)				WTP amount (in TZ shillings)			
	Maize	Sesame	Rice	Cassava	Maize	Sesame	Rice	Cassava
V	0.01	0.01	0.03	0.01	341.68	-114.60	-95.46	-162.15
	(0.03)	(0.06)	(0.04)	(0.06)	(209.14)	(353.43)	(175.45)	(277.70)
R	-0.04	0.03	-0.05	-0.04	150.42	115.59	-15.18	66.84
	(0.05)	(0.06)	(0.05)	(0.07)	(307.74)	(432.56)	(294.85)	(488.38)
R+V	-0.05	-0.06	-0.00	-0.12**	372.09	-78.32	374.82	-113.07
	(0.03)	(0.05)	(0.03)	(0.05)	(251.98)	(299.82)	(265.89)	(383.76)
Control mean	0.89	0.73	0.81	0.66	3413.3	2594.86	3000	2106.99
(Std. dev.)	(0.31)	(0.44)	(0.39)	(0.48)	(2732.78)	(2725.01)	(2806.64)	(2466.11)
N	959	727	806	549	959	727	806	549
R-squared	0.004	0.004	0.005	0.010	0.002	0.001	0.004	0.001
Village FE	YES	YES	YES	YES	YES	YES	YES	YES

Notes: V denotes Voucher group, R denotes Recommendations group, and R+V denotes the Recommendations and Voucher group. Control mean refers to the mean of the control group post-intervention. Robust standard errors in parentheses. Standard errors are clustered at the village level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

The first four columns in Table 18 report results of WTP for maize, sesame, rice, and cassava as an indicator variable, while the last four columns report WTP in TZ Shillings. As can be seen in the table, all of the coefficients are indistinguishable from zero, with an exception for the R+V farmers who have a 12% lower probability of desiring to pay for cassava. Overall, it seems that farmers are not interested in paying for soil recommendations, supporting the argument of farmers not being convinced of the R+V treatment's profitability. Overall, farmers' beliefs about their soil quality and/or the impact of fertilizer on net benefits plays a role in explaining farmers' reluctance to pay for soil information and adopt fertilizer, supporting the productivity results that are based on satellite images, which reveal no or minor treatment impact.

7 Conclusion

In this study, we examine the longer-term effects of alleviating credit and information constraints on fertilizer adoption and maize productivity. The results indicate a persistent treatment impact on maize SR yields obtained from the 2016-19 MMPs for farmers who were given both the recommendations and vouchers, but fertilizer use is indistinguishable from zero as early as one year after the intervention concluded. Employing more objective measures of productivity – such as by using satellite and GPS – reveals a lower and less significant treatment impact, supporting the argument that farmers do not view fertilizer as an investment that they are willing to engage in.

A more detailed analysis of other factors that may impede fertilizer adoption shows that, first, farmers who are informed about which fertilizers to apply and have the opportunity to act on and experiment with the information retain information in the longer run. Second, farmers update their prior perceptions upwards after receiving vouchers but not when they are given recommendations, suggesting that farmers are unaware of their soils' deficiencies. Third, assessment of willingness to pay indicates that farmers in all of the treatment groups are not willing to pay for soil information.

Identifying other possible mechanisms that affect fertilizer adoption and maize yields requires information on a variety of inputs. For instance, having data on organic fertilizer would allow testing whether some farmers applied more organic fertilizer that replenished their soils with carbon, an important nutrient that complements the nitrogen from fertilizer (Marenya & Barrett, 2009a, 2009b), and hence had higher productivity. Indeed, farmers may use inputs as complements. Research by BenYishay and Mobarak (2018) shows that farmers who adopted pit planting also had a 25% higher chance of applying manure. Composting is another technology that might have affected yields of the R+V farmers. This technology was shown in the same study by BenYishay and Mobarak (2018) to increase maize yields in Malawi by 50% when incentives and proximity were exploited to spread information about it. Moreover, in order to assess measurement error in output, full-plot crop cutting is needed instead of using SR data.

Farmers who applied fertilizer constitute a minority: 5.2%, 2.2% and 1.4% in 2017, 2018 and 2019 of farmers in our sample applied fertilizer on their 2014 MMP, respectively, compared to 0.8% in 2014. The proportions are not very different for the 2017-19 MMP: 4.9% in 2017, 2.4% in 2018, and 1.6% in 2019. These percentages highlight a persistent and an alarming lack of fertilizer use in Tanzania and raise concerns about many of the current programs that find short-term impacts on technology adoption and productivity. Farmers who apply fertilizer over the short-term only may in fact be learning that fertilizer use is unprofitable because their soils are so degraded, that repeated fertilizer application may be needed to increase profits substantially and change farmers' prior beliefs about their soil fertility.

Our findings also provide important insights for extension services. Dissemination of agronomic information to farmers without giving them the chance to experiment with the information may not result in long-term retention. Thus, to save costs of repeated extension visits, especially to remote villages, extension agents should consider finding ways of actively engaging farmers in exploring the information given to them rather than communicating with them as passive recipients only. Indeed, there has been a push lately towards increasing demonstration plots whereby farmers can observe the gains to improved agronomic practices like using mineral fertilizers. Timing of the information and the subsidies to act on it is also crucial because farmers retain information only when both fertilizer recommendations and vouchers are given at the same time, prior to the planting season when farmers are able to make use of the information.

More research is needed to decipher the mechanisms through which farmers update (or do not) their prior beliefs about their soil fertility. Understanding how soil perceptions are shaped is important because even if farmers have access to credit and are aware of the profit-maximizing fertilizer(s) types and quantities, they might choose to avoid applying fertilizers on soils that they regard as fertile, although these subjective beliefs are not necessarily true. When we understand how farmers' beliefs are shaped, it is possible to use behavioral insights, for example, to correct existing biases in case subjective perceptions do not match objective measures of soil quality. The findings of this study are a first step in this direction, but more studies are needed to deepen our understanding about the roles of market conditions, farmers' socioeconomic characteristics, and soils' biophysical properties in shaping farmers' perceptions about their soil fertility. Future research is also needed to determine how accurate are farmer reports of the area planted and yields using large scale interventions that quantify the factors leading to the divergence between GPS and SR areas, and satellite and SR productivity.

References

Abay, K. A., Abate, G. T., Barrett, C. B., & Bernard, T. (2019). Correlated non-classical measurement errors, 'second best' policy inference, and the inverse size-productivity relationship in agriculture. *Journal of Development Economics*, *139*, 171 - 184. doi: https://doi.org/10.1016/j.jdeveco.2019.03.008

Abay, K. A., Bevis, L. E. M., & Barrett, C. B. (2020). Measurement error mechanisms matter: Agricultural intensification with farmer misperceptions and misreporting. *American Journal of Agricultural Economics*, *n/a*(n/a). doi: https://doi.org/10.1111/ajae.12173

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Bandiera, O., & Rasul, I. (2006). Social networks and technology adoption in Northern Mozambique. *Economic Journal*, *116*(514), 869-902. doi: 10.1111/j.1468-0297.2006.01115.x

Beaman, L., BenYishay, A., Magruder, J., & Mobarak, A. M. (2018). *Can network theory-based targeting increase technology adoption?* (Working Paper No. 24912). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w24912 doi: 10.3386/w24912

BenYishay, A., & Mobarak, A. M. (2018). Social learning and incentives for experimentation and communication. *The Review of Economic Studies*, *86*(3), 976-1009. doi: 10.1093/restud/rdy039

Berazneva, J., McBride, L., Sheahan, M., & Güereña, D. (2018). Empirical assessment of subjective and objective soil fertility metrics in east Africa: Implications for researchers and policy makers. *World Development*, *105*, 367-382. doi: https://doi.org/10.1016/j.worlddev.2017.12.009

Burke, M., & Lobell, D. B. (2017). Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1616919114

Carletto, C., Gourlay, S., & Winters, P. (2015). From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis. *Journal of African Economies*, *24*(5), 593-628. doi: 10.1093/jae/ejv011

Carletto, C., Savastano, S., & Zezza, A. (2013). Fact or artifact: The impact of measurement errors on the farm size–productivity relationship. *Journal of Development Economics*, *103*, 254 - 261. doi: https://doi.org/10.1016/j.jdeveco.2013.03.004

Conley, T. G., & Udry, C. R. (2010). Learning about a new technology: Pineapple in Ghana. *American Economic Review*, *100*(1), 35-69. doi: 10.1257/aer.100.1.35

Croppenstedt, A., Demeke, M., & Meschi, M. M. (2003). Technology adoption in the presence of constraints: The case of fertilizer demand in Ethiopia. *Review of Development Economics*, *7*(1), 58–70. doi: https://doi.org/10.1111/1467-9361.00175

Duflo, E., Kremer, M., & Robinson, J. (2008). How high are Rates of return to fertilizer? Evidence from field experiments in Kenya. *American Economic Review*, *98*(2), 482-88. doi: 10.1257/aer.98.2.482

Duflo, E., Kremer, M., & Robinson, J. (2011). Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya. *American Economic Review*, *101*(6), 2350-90. Retrieved from https://www.aeaweb.org/

articles?id=10.1257/aer.101.6.2350 doi: 10.1257/aer.101.6.2350

Emerick, K., de Janvry, A., Sadoulet, E., & Dar, M. H. (2016). Technological innovations, downside risk, and the modernization of agriculture. *American Economic Review*, *106*(6), 1537-61. doi: 10.1257/aer.20150474

Gourlay, S., Kilic, T., & Lobell, D. B. (2019). A new spin on an old debate: Errors in farmer-reported production and their implications for inverse scale - productivity relationship in Uganda. *Journal of Development Economics*, *141*, 102376. doi: https://doi.org/10.1016/j.jdeveco.2019.102376

Hanemann, M., Loomis, J., & Kanninen, B. (1991). Statistical efficiency of double-bounded dichotomous choice contingent valuation. *American Journal of Agricultural Economics*, *73*(4), 1255-1263. doi: https://doi.org/10.2307/1242453

Harou, A. P., Madajewicz, M., Michelson, H., Palm, C. A., Amuri, N., Magomba, C., … Weil, R. (2020). The joint effects of information and financing constraints on technology adoption: Evidence from a field experiment in rural Tanzania. *Submitted*.

Harou, A. P., Walker, T. F., & Barrett, C. B. (2017). Is late really better than never? The farmer welfare effects of pineapple adoption in Ghana. *Agricultural Economics*, *48*(2), 153-164. doi: 10.1111/agec.12322

Haushofer, J., & Shapiro, J. (2016). The Short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya. *The Quarterly Journal of Economics*, *131*(4), 1973-2042. doi: 10.1093/qje/qjw025

Jama, B., Kimani, D., Harawa, R., Mavuthu, A. K., & Sileshi, G. W. (2017). Maize yield response, nitrogen use efficiency and financial returns to fertilizer on smallholder farms in southern Africa. *Food Security*, *9*, 577–593. doi: 10.1007/s12571-017-0674-2

Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., & Lobell, D. B. (2019). Smallholder maize area and yield mapping at national scales with google earth engine. *Remote Sensing of Environment*, *228*, 115 - 128. doi: https://doi.org/10.1016/j.rse.2019.04.016

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*, *374*(2065). doi: https://doi.org/10.1098/rsta.2015.0202

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141-151. doi: 10.1177/001316446002000116

Kaizzi, K., Byalebeka, J., Semalulu, O., Alou, I. N., Zimwanguyizza, W., Nansamba, A., … Wortmann, C. (2012). Maize response to fertilizer and nitrogen use efficiency in Uganda. *Agronomy Journal*, *104*(1), 73-82. doi: 10.2134/agronj2011.0181

Linn, R. L. (1968). A Monte Carlo approach to the number of factors problem. *Psychometrika*, *33*(1), 37-71. doi: https://doi.org/10.1007/bf02289675

Lobell, D. B., Azzari, G., Burke, M., Gourlay, S., Jin, Z., Kilic, T., & Murray, S. (2019). Eyes in the sky, boots on the ground: Assessing satellite- and ground-based approaches to crop yield measurement and analysis. *American Journal of Agricultural Economics*. doi: 10.1093/ajae/aaz051

Marenya, P. P., & Barrett, C. B. (2009a). Soil quality and fertilizer use rates among smallholder farmers in western Kenya. *Agricultural Economics*, *40*(5), 561–572. doi: 10.1111/j.1574-0862.2009.00398.x

Marenya, P. P., & Barrett, C. B. (2009b). State-Conditional fertilizer yield response on Western Kenyan farms. *American Journal of Agricultural Economics*, *91*(4), 991-1006. doi: 10.1111/j.1467-8276.2009.01313.x

Marenya, P. P., Barrett, C. B., & Gulick, T. (2008). *Farmers' perceptions of soil fertility and fertilizer yield response in Kenya* (Working Paper). Social Science Research Network. Retrieved from http://dx.doi.org/10.2139/ssrn.1845546

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more t in experiments. *Journal of Development Economics*, *99*(2), 210 - 221. doi: https://doi.org/10.1016/j.jdeveco.2012.01.002

Michelson, H., Fairbairn, A., Ellison, B., Maertens, A., & Manyong, V. (2021). Misperceived quality: Fertilizer in tanzania. *Journal of Development Economics*, *148*, 102579. doi: https://doi.org/10.1016/j.jdeveco.2020.102579

Mowo, J., Floor, J., Kaihura, F., & Magoggo, J. (1993). *Review of fertilizer recommendations in Tanzania: Part 2, revised fertilizer recommendations for Tanzania* (No. Soil Fertility Report No. F6). Retrieved from https://library.wur.nl/isric/fulltext/isricu_i13557_001.pdf ([Online; accessed May 8, 2020])

Mtaki, B. (2017). *2017 Tanzania corn, wheat and rice report.* Retrieved from https://apps.fas.usda.gov/newgainapi/api/report/downloadreportbyfilename?filename=Grain%20and%20Feed%20Annual_Dar%20es%20Salaam_Tanzania%20-%20United%20Republic%20of_4-6-2018.pdf ([Online; accessed April 5, 2020])

Mukherjee, A., & Lal, R. (2014). Comparison of soil quality index using three methods. *PLoS ONE*, *9*(8). doi: 10.1371/journal.pone.0105981

Murphy, D. M., Roobroeck, D., Lee, D. R., & Thies, J. (2020). Underground knowledge: Estimating the impacts of soil information on transfers through experimental auctions. *American Journal of Agricultural Economics*, *00*(00), 1-26. doi: 10.1111/ajae.12101

Paavola, J. (2008). Livelihoods, vulnerability and adaptation to climate change in Morogoro, Tanzania. *Environmental Science Policy*, *11*(7), 642-654. doi: https://doi.org/10.1016/j.envsci.2008.06.002

Rowe, E., van Wijk, M., de Ridder, N., & Giller, K. (2006). Nutrient allocation strategies across a simplified heterogeneous African smallholder farm. *Agriculture, Ecosystems and Environment*, *116*(1-2), 60–71.

Shee, A., Azzarri, C., & Haile, B. (2020). Farmers' willingness to pay for improved agricultural technologies: Evidence from a field experiment in tanzania. *Sustainability*, *12*(1), 216. doi: https://doi.org/10.3390/su12010216

Suri, T. (2011). Selection and comparative advantage in technology adoption. *Econometrica*, *79*(1), 159-209. doi: 10.3982/ECTA7749

Tittonell, P., Muriuki, A., Klapwijk, C. J., Shepherd, K. D., Coe, R., & Vanlauwe, B. (2013). Soil heterogeneity and soil fertility gradients in smallholder farms of the East African Highlands. *Soil Science Society of America Journal*, *77*(2), 525-538. doi: 10.2136/sssaj2012.0250

Vanlauwe, B., Kihara, J., Chivenge, P., Pypers, P., Coe, R., & Six, J. (2011). Agronomic use efficiency of N fertilizer in maize-based systems in sub-Saharan Africa within the context of integrated soil fertility management. *Plant* *and Soil*, *339*, 35-50. doi: 10.1007/s11104-010-0462-7

Xu, Z., Guan, Z., Jayne, T. S., & Black, R. (2009). Factors influencing the profitability of fertilizer use on maize in Zambia. *Agricultural Economics*, *40*(4), 437–446. doi: 10.1111/j.1574-0862.2009.00384.x

Zingore, S., & Johnston, A. (2013). The 4R nutrient stewardship in the context of smallholder agriculture in Africa. In B. Vanlauwe, P. van Asten, & G. Blomme (Eds.), *Agro-Ecological intensification of agricultural systems in the African Highlands* (p. 77-84). London: Routledge.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432-442. doi: https://psycnet.apa.org/doi/10.1037/0033-2909.99.3.432

A Appendices

A.1 Attrition Analysis

First, we check that the probability of attriting is not correlated with treatment by estimating the following equation using ordinary least squares (OLS) with village fixed effects (FE):

$$attrition_{iv} = \alpha_0 + \sum_{k=1}^{3} \theta_k TREAT_i^k + d_v + \varepsilon_{iv} \tag{22}$$

where $attrition_{iv}$ is a binary variable that takes the value one if farmer i in village v attrited and 0 otherwise, α_0 is a constant, $TREAT_i^k$ is a binary variable that takes the value one for each farmer i assigned to one of the k treatment arms (V, R and R+V) and zero otherwise, d_v is village FE, and ε_{iv} is the associated idiosyncratic error term. The omitted category is the control group, and standard errors are clustered at the village-level. Table 20 indicates that there is no statistical difference in probability of attrition between all the treatments compared to the control group.

Table 19: Probability of Attrition by Treatment

	attrition (2016)		attrition (2019)	
V	-0.00558	-0.00326	-0.0147	-0.0314
	(0.0241)	(0.0214)	(0.0297)	(0.0295)
R	0.0122	0.0157	0.0361	0.0201
	(0.0239)	(0.0217)	(0.0301)	(0.0327)
R+V	0.00290	0.00539	0.0222	0.00423
	(0.0188)	(0.0178)	(0.0281)	(0.0260)
Constant	0.0611***		0.116***	
	(0.0135)		(0.0139)	
N	1,050	1,050	1,050	1,050
R-squared	0.001	0.001	0.003	0.003
Village FE	NO	YES	NO	YES

Notes: V denotes the Voucher group, R denotes the Recommendations group, and R+V denotes the Recommendations and Voucher group. Robust standard errors in parentheses. Standard errors are clustered at the village level.
*** p<0.01, ** p<0.05, * p<0.1

Second, we test whether attrition is associated with any of the outcome variables of interest, which are SSB, fertilizers and yields in kg/acre, and fertilizer use as a binary variable, by running the following OLS regression with village FE:

$$y_{iv} = \alpha_0 + attrition_i + d_v + \varepsilon_{iv} \tag{23}$$

where y_{iv} is one of the aforementioned outcome variables, and the remaining variables are the same as in equation (22). The regression results are reported in Table 21. We do not include one of the outcome variables, namely the retention index, because it is analyzed following cross-sectional estimations in 2016 and then in 2019 (post-intervention), as we ask farmers to recall the recommendations given to them at baseline.

Table 21 confirms that attrition did not have a statistical impact on the estimates of the main outcome variables, except when fertilizer use is an indicator variable. A total of eight farmers applied fertilizer among the 1,050 participating farmers at baseline, and all of them attrited from the study. We believe this might be the reason driving the statistical significance. Nonetheless, the coefficient does not seem to be economically significant since those who attrited have a 0.7% and 1.1% less chance of applying fertilizer in 2016 and 2019, respectively.

Table 20: Attrition Effect on Outcome Variables

	Year	Fertilizer (kg/acre)	Fertilizer (=1)	Yields (kg/ SR acre)	Yields (kg/ GPS acre)	SSB=0 poor; 1 fair; 2 good
attrition (=1)	2016	-0.20	-0.01**	-29.75	-58.25	0.03
		(0.17)	(0.00)	(40.59)	(67.89)	(0.07)
attrition (=1)	2019	-0.23	-0.01***	-12.22	-	0.08
		(0.15)	(0.00)	(34.68)		(0.05)
N		1,050	1,050	915	752	1,046
Village FE		YES	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors are clustered at the village level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

A.2 Description of Baseline Variables

Table 21: Description of Baseline Variables

Variable	Description
Dependency	Dependency ratio is the proportion of individuals aged less than 18 and over 60 years
Assets	Asset index is constructed from quantities of household, productive, and livestock assets using principal component analysis (PCA)
Age	Age of household head in years
Gender	=1 if household head is male and 0 if female
Educ	=1 if household head completed any education and 0 otherwise
Educ2	=1 if household head completed more than primary education and 0 otherwise
Credit	=1 if household accessed credit and 0 otherwise
Distance	Distance to the 2014 main maize plot in minutes
Total Area	Total area owned by the household in acres
Maize Area	Total area planted with maize by the household in acres
Seeds	=1 if household used improved seeds and 0 otherwise

A.3 Robustness to Pooling

Table 22: Pooled Treatment Effect on Self-Reported Adoption and Productivity

	2014 MMP			2016-19 MMP		
	Fertilizer (kg/acre)	Fertilizer (=1)	Yields (kg/acre)	Fertilizer (kg/acre)	Fertilizer (=1)	Yields (kg/acre)
2016	1.03***	0.04***	-148.43***	1.23***	0.04***	-144.26***
	(0.34)	(0.01)	(28.97)	(0.35)	(0.01)	(28.72)
2017	0.80***	0.02**	53.46**	0.72**	0.02**	55.39**
	(0.30)	(0.01)	(22.53)	(0.28)	(0.01)	(21.34)
2018	0.23	0.01	-5.74	0.27	0.01	5.30
	(0.17)	(0.01)	(25.32)	(0.18)	(0.01)	(22.07)
2019	0.47	0.01	-124.32***	0.45*	0.01	-120.89***
	(0.29)	(0.01)	(31.33)	(0.23)	(0.01)	(26.65)
TREAT	-0.00	-0.00	-43.19*	0.02	-0.00	-50.93**
	(0.24)	(0.01)	(24.81)	(0.23)	(0.01)	(24.40)
2016×TREAT	16.73***	0.50***	118.57***	16.93***	0.51***	111.24***
	(1.59)	(0.04)	(43.44)	(1.65)	(0.04)	(40.46)
2017×TREAT	1.51**	0.05***	90.82**	1.64**	0.05***	116.84***
	(0.58)	(0.02)	(39.07)	(0.66)	(0.02)	(36.63)
2018×TREAT	0.38	0.02	100.87**	1.03	0.03**	127.39***
	(0.61)	(0.01)	(41.07)	(0.78)	(0.01)	(44.81)
2019×TREAT	0.23	-0.00	72.67*	0.31	-0.00	108.72**
	(0.84)	(0.01)	(40.73)	(0.72)	(0.01)	(41.75)
Control mean	0.09	0.01	441.26	0.09	0.01	441.26
(std. dev.)	(1.98)	(0.09)	(368.63)	(1.98)	(0.09)	(368.63)
N	3,434	3,456	3,273	4,123	4,144	4,002
R-squared	0.192	0.340	0.043	0.184	0.330	0.049
Village FE	YES	YES	YES	YES	YES	YES

Notes: TREAT is equal to unity if a farmer was assigned to the V or R+V groups and is equal to zero if a farmer was assigned the C or R groups. Robust standard errors in parentheses. Standard errors are clustered at the village level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 23: Pooled Treatment Effect on Self-Reported and Satellite-Derived Productivity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	SR	Satellite	SR	Satellite	SR	Satellite	SR	Satellite
	2016	2016	2019	2019	2016	2016	2019	2019
Panel A: 2014 MMP								
TREAT (=1)	58.68	0.07**	0.79	-0.01	86.46***	0.03	39.32*	-0.03
	(42.29)	(0.03)	(36.05)	(0.04)	(32.09)	(0.02)	(22.05)	(0.04)
Pure stand (=1)	-35.47	-0.08	34.17	-0.02	-75.75	-0.06	52.76	-0.02
	(51.06)	(0.05)	(40.37)	(0.04)	(46.80)	(0.06)	(46.58)	(0.04)
Control mean	473.55	2.18	444.39	2.45	473.55	2.18	444.39	2.45
(Std. dev.)	(377.91)	(0.41)	(321.29)	(0.41)	(377.91)	(0.41)	(321.29)	(0.41)
N	659	659	409	409	659	659	409	409
R-squared	0.005	0.008	0.002	0.000	0.013	0.003	0.007	0.002
Village FE	NO	NO	NO	NO	YES	YES	YES	YES
Panel B: 2016/19 MMP								
TREAT (=1)	56.62	0.04	50.76	0.05	88.78***	0.02	48.04**	0.01
	(39.54)	(0.03)	(33.61)	(0.04)	(30.16)	(0.03)	(23.03)	(0.03)
Pure stand (=1)	-20.05	-0.07	74.79**	0.02	-67.86	-0.05	74.63**	0.03
	(52.60)	(0.05)	(29.64)	(0.04)	(49.33)	(0.06)	(32.58)	(0.03)
Control mean	467.91	2.20	445.65	2.42	467.91	2.20	445.65	2.42
(Std. dev.)	(375.86)	(0.42)	(332.42)	(0.41)	(375.86)	(0.42)	(332.42)	(0.41)
N	699	699	689	689	699	699	689	689
R-squared	0.008	0.001	0.006	0.000	0.007	0.001	0.007	0.000
Village FE	NO	NO	NO	NO	YES	YES	YES	YES

Notes: TREAT is equal to unity if a farmer was assigned to the V or R+V groups and is equal to zero if a farmer was assigned the C or R groups. Robust standard errors in parentheses. Standard errors are clustered at the village level.
*** p<0.01, ** p<0.05, * p<0.1

Table 24: Pooled Treatment Effect on Productivity Using Self-Reported and GPS Plot Size

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2014 MMP		>0.5 acres		2016/19 MMP		>0.5 acres	
	SR	GPS	SR	GPS	SR	GPS	SR	GPS
2016	-150.34***	-127.50***	-150.06***	-138.79***	-154.98***	-126.12***	-154.83***	-143.58***
	(30.34)	(31.99)	(31.66)	(29.88)	(28.87)	(31.30)	(30.14)	(28.88)
2019	-125.58***	-104.67***	-135.00***	-109.21***	-130.66***	-84.05***	-139.63***	-91.86***
	(31.74)	(30.14)	(30.26)	(29.04)	(25.82)	(24.59)	(25.81)	(23.87)
TREAT	-51.12**	-19.90	-52.10**	-29.81	-61.76***	-27.30	-63.63***	-38.25
	(22.75)	(25.93)	(23.39)	(26.60)	(22.81)	(25.82)	(23.40)	(26.55)
2016×TREAT	133.84***	61.13	135.82**	75.61*	134.78***	59.90	136.90***	81.30*
	(49.36)	(44.51)	(50.93)	(44.26)	(46.47)	(43.33)	(47.94)	(42.76)
2019×TREAT	84.04**	76.25	90.87**	75.83	131.35***	100.51*	138.13***	97.69**
	(39.91)	(52.20)	(36.83)	(48.42)	(38.19)	(50.19)	(38.14)	(41.38)
Control mean	450.02	406.59	453.58	394.37	450.02	406.59	453.58	394.37
(Std. dev.)	(367.97)	(391.01)	(373.59)	(376.21)	(367.97)	(391.01)	(373.59)	(376.21)
N	1,817	1,817	1,700	1,700	2,137	2,137	1,981	1,981
R-squared	0.024	0.019	0.026	0.024	0.025	0.016	0.027	0.023
Village FE	YES	YES	YES	YES	YES	YES	YES	YES

Notes: TREAT is equal to unity if a farmer was assigned to the V or R+V groups and is equal to zero if a farmer was assigned the C or R groups. Robust standard errors in parentheses. Standard errors are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

A.4 Robustness to ANCOVA Estimation

Table 25: Treatment Effect on Self-Reported Adoption and Productivity (ANCOVA)

	2014 MMP			2016-19 MMP		
	Fertilizer (kg/acre)	Fertilizer (=1)	Yields (kg/acre)	Fertilizer (kg/acre)	Fertilizer (=1)	Yields (kg/acre)
2016						
V	8.99***	0.31***	49.71	9.30***	0.32***	43.59
	(2.33)	(0.07)	(41.61)	(2.27)	(0.06)	(41.18)
R	1.05	0.04*	23.81	1.16	0.04*	15.80
	(1.07)	(0.02)	(31.49)	(1.03)	(0.02)	(32.17)
R+V	24.90***	0.71***	122.93***	25.03***	0.71***	105.22***
	(2.34)	(0.03)	(40.63)	(2.18)	(0.03)	(38.00)
Baseline value	0.72**	0.14	0.11**	0.71**	0.15	0.11**
	(0.30)	(0.17)	(0.05)	(0.30)	(0.17)	(0.05)
Control mean	0.03	0.01	468.45	0.03	0.01	459.89
(Std. dev.)	(0.33)	(0.09)	(401.05)	(0.32)	(0.09)	(396.08)
N	848	870	822	874	895	888
R-squared	0.209	0.371	0.022	0.209	0.374	0.019
Village FE	YES	YES	YES	YES	YES	YES
2017						
V	1.39	0.03	8.96	1.86	0.04	35.62
	(1.29)	(0.04)	(54.01)	(1.14)	(0.03)	(46.75)
R	-1.04	-0.04	-26.31	-0.60	-0.02	-32.00
	(0.88)	(0.03)	(62.74)	(0.66)	(0.02)	(56.68)
R+V	0.80	0.05	58.89	1.06	0.06	60.01
	(0.99)	(0.04)	(55.18)	(0.95)	(0.03)	(47.78)
Baseline value	0.02**	0.13	0.24***	-0.03***	0.06	0.24***
	(0.01)	(0.15)	(0.08)	(0.01)	(0.10)	(0.07)
Control mean	0.04	0.01	465.60	0.03	0.01	456.12
(Std. dev.)	(0.38)	(0.10)	(398.59)	(0.34)	(0.09)	(381.36)
N	621	621	621	782	782	782
R-squared	0.009	0.018	0.039	0.010	0.015	0.043
Village FE	YES	YES	YES	YES	YES	YES
2018						
V	1.26	0.03	10.74	1.59	0.04**	20.26
	(1.00)	(0.02)	(73.51)	(0.99)	(0.02)	(65.54)
R	-0.31	-0.01	-110.85	0.24	-0.00	-68.11
	(0.20)	(0.01)	(82.47)	(0.40)	(0.01)	(60.98)
R+V	-0.09	0.01	-16.26	1.05	0.03*	55.11
	(0.21)	(0.02)	(70.31)	(0.96)	(0.02)	(49.83)
Baseline value	-0.00	-0.04	0.19**	-0.04**	-0.03**	0.20***
	(0.02)	(0.05)	(0.08)	(0.02)	(0.02)	(0.07)
Control mean	0.04	0.01	480.90	0.03	0.01	467.07
(Std. dev.)	(0.39)	(0.09)	(373.24)	(0.35)	(0.10)	(352.47)
N	499	499	499	710	710	710
R-squared	0.011	0.009	0.030	0.007	0.010	0.031
Village FE	YES	YES	YES	YES	YES	YES
2019						
V	0.28	-0.00	38.72	1.08	0.01	50.51
	(1.33)	(0.02)	(31.64)	(1.37)	(0.02)	(35.26)
R	-1.28	-0.01	16.37	-0.20	0.01	36.84
	(0.98)	(0.03)	(50.92)	(0.69)	(0.02)	(39.12)
R+V	-1.50*	-0.02	55.53	-0.59	-0.01	85.10**
	(0.89)	(0.02)	(36.16)	(0.43)	(0.01)	(35.72)
Baseline value	-0.08***	-0.06*	0.16*	-0.08**	-0.03	0.17***
	(0.02)	(0.03)	(0.09)	(0.03)	(0.02)	(0.05)
Control mean	0.04	0.01	451.21	0.03	0.01	453.23
(Std. dev.)	(0.42)	(0.10)	(346.38)	(0.35)	(0.10)	(334.41)
N	416	416	416	707	707	707
R-squared	0.011	0.007	0.025	0.008	0.003	0.030
Village FE	YES	YES	YES	YES	YES	YES
All: 2016-2019						
V	4.01***	0.13***	32.66	3.72***	0.11***	37.57
	(1.28)	(0.03)	(39.00)	(1.21)	(0.03)	(36.69)
R	-0.26	-0.00	-14.39	0.18	0.01	-9.01
	(0.51)	(0.01)	(43.65)	(0.46)	(0.01)	(36.37)
R+V	8.97***	0.28***	67.30*	7.73***	0.23***	79.39**
	(1.05)	(0.02)	(36.93)	(0.89)	(0.02)	(32.34)
Baseline value	0.33***	0.11	0.18***	0.16**	0.04	0.18***
	(0.10)	(0.07)	(0.05)	(0.07)	(0.05)	(0.04)
Control mean	0.02	0.01	448.61	0.02	0.007	448.61
(Std. dev.)	(0.30)	(0.08)	(386.97)	(0.30)	(0.08)	(386.97)
N	2,384	2,406	2,358	3,073	3,094	3,087
R-squared	0.116	0.215	0.070	0.107	0.202	0.077
Village FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

Notes: V denotes Voucher group, R denotes the Recommendations group, and RV denotes the Recommendations and Voucher group. Regressions include indicators for missing values following equation (3). Robust standard errors in parentheses. Standard errors are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 26: Treatment Effect on Productivity Using Self-Reported and GPS Plot Size (ANCOVA)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2014 MMP		>0.5 acres		2016/19 MMP		>0.5 acres	
	SR	GPS	SR	GPS	SR	GPS	SR	GPS
2016								
V	49.75	30.06	43.33	33.19	53.74	39.24	47.64	40.22
	(46.62)	(42.40)	(48.19)	(44.16)	(46.43)	(40.00)	(48.47)	(41.36)
R	-11.68	-23.01	-17.14	-29.38	-21.43	-21.03	-26.52	-26.54
	(30.15)	(30.70)	(31.32)	(29.87)	(30.55)	(28.69)	(34.42)	(28.20)
R+V	117.71**	53.51	119.19**	54.90	111.42***	59.58	112.31**	59.86
	(45.03)	(49.72)	(46.06)	(53.29)	(40.86)	(49.37)	(42.50)	(52.17)
Baseline value	0.13**	0.23***	0.12*	0.22***	0.12**	0.23***	0.11*	0.23***
	(0.06)	(0.07)	(0.06)	(0.07)	(0.05)	(0.07)	(0.06)	(0.07)
Control mean	481.13	435.74	485.28	427.91	474.38	429.30	480.90	456.84
(Std. dev.)	(399.58)	(411.28)	(405.52)	(409.08)	(395.66)	(411.59)	(402.43)	(337.21)
N	659	659	620	620	699	699	656	656
R-squared	0.024	0.044	0.024	0.048	0.023	0.048	0.023	0.054
Village FE	YES	YES	YES	YES	YES	YES	YES	YES
2019								
V	43.85	45.28	13.62	17.09	51.71	48.29	37.49	43.27
	(34.30)	(40.14)	(35.85)	(38.16)	(36.37)	(36.48)	(37.33)	(35.92)
R	10.10	-29.97	-25.70	-56.45	30.55	26.43	16.12	13.50
	(50.03)	(46.55)	(51.16)	(49.93)	(39.03)	(51.42)	(42.88)	(49.37)
R+V	52.59	61.33	28.22	12.73	87.76**	94.87*	75.02**	53.27
	(38.41)	(46.32)	(34.59)	(41.59)	(36.15)	(54.50)	(34.43)	(38.56)
Baseline value	0.16*	0.19***	0.11	0.16***	0.17***	0.14**	0.14***	0.14***
	(0.09)	(0.07)	(0.07)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Control mean	456.39	431.86	454.14	408.48	456.84	418.48	461.46	407.49
(Std. dev.)	(347.30)	(392.28)	(353.55)	(345.40)	(337.21)	(374.62)	(343.87)	(346.63)
N	409	409	372	372	689	689	617	617
R-squared	0.025	0.043	0.012	0.035	0.031	0.021	0.025	0.029
Village FE	YES	YES	YES	YES	YES	YES	YES	YES
All: 2016 & 2019								
V	55.52	42.43	41.03	33.96	52.74	41.14	42.27	40.09
	(35.49)	(31.54)	(35.62)	(32.37)	(33.88)	(29.96)	(34.20)	(29.86)
R	8.23	-12.54	-9.15	-27.26	7.77	2.78	-3.57	-8.65
	(30.82)	(28.24)	(29.98)	(25.18)	(31.68)	(36.01)	(33.52)	(32.77)
R+V	92.05**	55.33	82.03**	37.47	99.23***	72.87*	91.26***	50.87
	(36.01)	(41.25)	(35.91)	(39.05)	(30.73)	(41.82)	(30.78)	(37.68)
Baseline value	0.15***	0.22***	0.13**	0.22***	0.14***	0.18***	0.12***	0.19***
	(0.05)	(0.06)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)
Control mean	455.23	417.92	456.98	402.67	455.23	417.92	456.98	402.67
(Std. dev.)	(385.24)	(414.67)	(390.47)	(394.80)	(385.24)	(414.67)	(390.47)	(394.80)
N	1,068	1,068	992	992	1,388	1,388	1,273	1,273
R-squared	0.020	0.044	0.018	0.047	0.022	0.035	0.020	0.044
Village FE	YES	YES	YES	YES	YES	YES	YES	YES

Notes: V denotes Voucher group, R denotes the Recommendations group, and RV denotes the Recommendations and Voucher group. Regressions include indicators for missing values following equation (3). Robust standard errors in parentheses. Standard errors are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 27: Summary of Results

	Data	DD	OLS	Pooled	ANCOVA
Panel A: 2014 MMP					
2016	SR	37%***	NA	27%***	26%***
2016	Satellite	NA	0-4%*	0-3%**	NA
2016	GPS	0%	NA	0-19%*	0%
2019	SR	0%	NA	17%*	0%
2019	Satellite	NA	0%	0%	NA
2019	GPS	0%	NA	0%	0%
Panel B: 2016/19 MMP					
2016	SR	34%***	NA	25%***	23%***
2016	Satellite	NA	0%	0%	NA
2016	GPS	0%	NA	0-21%*	0%
2019	SR	33%***	NA	25%**	19%**
2019	Satellite	NA	0-3.5%**	0%	NA
2019	GPS	33%*	NA	25%**	23%*

Notes: DD denotes difference-in-differences and Pooled denotes the pooled V and R+V groups. *** p<0.01, ** p<0.05, * p<0.1

A.5 Construction of the True Soil Quality Index

The true soil quality (TSQ) index is constructed as follows. First, we convert the individual soil nutrient measures to a number between zero and one. For pH and electrical conductivity (EC), we use an optimal scoring method. For example, the optimal pH value is between 5.5 and 6.5, which is assigned a value of one. As the pH decreases or increases away from this optimal value, the score decreases progressively. For the other nutrient and/or soil characteristics, a 'more is better' scoring method is applied. The scores attributed to the level of nutrients are shown in Table **??** below. Second, once the individual nutrients have been assigned a value between zero and one, with a higher number

Table 28: Individual Nutrient Scores

pH (H2O)							
Score	0	0.33	0.66	1	0.66	0.33	0
Classification	Very low	Low	Medium	Optimum	High-not limiting	High-limiting	Very high
Thresholds	<= 4	>4 & <= 5	>5 & <= 5.5	>5.5 & <=6.5	>6.5 & <=7.5	>7.5 & <=8.5	>8.5
Electrical conductivity							
Score	0	0.33	0.66	1	0.66	0.33	0
Classification	Low fertility	Medium fertility	Slightly saline	Very saline	Severe salinity	Very severe salinity	Few crops can grow
Thresholds	<= 0.1	>0.1 & <= 0.3	>0.3 & <= 0.6	>0.6 & <=1.2	>1.2 & <=2.4	>2.4 & <=4	>4
Phosphorus							
Score	0	0.33	0.33	0.66	0.66	1	
Classification	Extremely low	Very low	Low	Medium	High	Excessive	
Thresholds	<= 0.05	>0.05 & <= 0.1	>0.1 & <= 0.3	>0.3 & <= 0.5	>0.5 & <= 2	>2	
Potassium							
Score	0	0.25	0.5	0.75	1		
Classification	Very low	Low	Medium	High	Very high		
Thresholds	<= 10	>10 & <= 20	>20 & <= 40	>40 & <= 60	>60		
Sulfur							
Score	0	0.33	0.66	1			
Classification	Very low	Low	Medium	High			
Thresholds	<= 5	>5 & <= 10	>10 & <= 15	>15			
Active carbon							
Score	0	0.2	0.4	0.6	0.8	1	
Classification	Extremely low	Very low	Low	Medium	High	Extremely high	
Thresholds	<= 150	>150 & <= 250	>250 & <= 350	>350 & <=500	>500 & <=700	>700	
Nitrogen							
Score	0	0.2	0.4	0.6	0.8	1	
Classification	Very low	Low	Medium	Medium/High	High	Very high	
Thresholds	<= 21	>21 & <= 42	>42 & <= 65	>65 & <=90	>90 & <=120	>120	

Notes: Electrical conductivity is measured in units of deciSiemens per meter (dS/m). Phosphorus is measured in units of mg P per kg of soil. Potassium is measured in units of mg K per kg of soil. Sulfur is measured in units of mg S per kg of soil. Active carbon is measured in units of mg carbon per kg of soil. Nitrogen is measured in units of mg N per kg of soil.

indicating a better nutritional content, we use Principal Component Analysis (PCA) to assign weights to every nutrient. PCA is a statistical method that is used to reduce data dimensions for ease of interpretation, but without losing the information explained (Jolliffe & Cadima, 2016). This is done by creating new variables that we call principal components. In the context of this study, we are assuming that the seven soil fertility indicators explain a common

Table 29: PCA Eigenvalues and Variances

	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7
Eigenvalues	1.55	1.33	1.14	0.96	0.77	0.65	0.60
Variance proportion	0.22	0.19	0.16	0.14	0.11	0.092	0.086
Cumulative variance	0.22	0.41	0.57	0.71	0.82	0.91	1

factor, which is soil fertility. By using PCA, we are able to reduce the seven soil fertility indicators to a number of principal components that can be used to obtain weights for every variable.

Different methods have been used to determine how many components should be retained when doing PCA. The most used method is to retain all components that have an eigenvalue larger than one, which is known as the Kaiser criterion (Kaiser, 1960). Thus, the first three components are retained as they have eigenvalues greater than one. However, several studies showed that the Kaiser rule can be inaccurate (Linn, 1968; Zwick & Velicer, 1986).

Since the total variance explained by these three components is only around 57% (Table **??**), we use the "cumulative percent of variance accounted for" criterion and retain the fourth component as well that has an eigenvalue of 0.96, raising the total variance explained to 71% (Table **??**). Keeping the fourth component is also intuitive due to two main reasons. First, the total variance explained is now more than 70%, which is a common minimum total variance threshold used (Jolliffe & Cadima, 2016). Second, the eigenvalue of this additional retained component is 0.96, which is very close to one. Then, once the four components are retained, each nutritional element is assigned a weight by the fraction of total variation explained by its component out of the total variance. The relevant component for each nutritional element is the one which corresponds to the maximum eigenvector observed in Table **??**.[10] The same procedure is applied to all other nutrient elements to assign a weight for each one of them. Once all weights have been assigned, the soil index is then calculated according to the following formula:

$$Soil\ Index = \sum_{i=1}^{7} Weight \times Score_i \tag{24}$$

where $Weight$ is the PCA weights explained above, i corresponds to each nutritional element, and Score is a value between zero and one of each nutritional element detailed above. To expand (24), we use the assigned weights:

$$\begin{aligned} Soil\ Index = \left(\frac{0.16}{0.71} \times pH\right) + \left(\frac{0.19}{0.71} \times EC\right) + \left(\frac{0.14}{0.71} \times C\right) + \left(\frac{0.19}{0.71} \times N\right) \\ + \left(\frac{0.22}{0.71} \times P\right) + \left(\frac{0.22}{0.71} \times K\right) + \left(\frac{0.22}{0.71} \times S\right) \end{aligned} \tag{25}$$

Finally, the index is then normalized to a number between 0 and 1 by adding the lowest score (in absolute value) and then dividing by the highest value to get the TSQ index. A TSQ is considered poor if it is less than 0.33; fair if it is higher than or equal to 0.33 and less than 0.66; good if it is higher than or equal to 0.66. Figure 10 shows the distribution of the index and the cutoff lines.

[10]For example, pH is assigned to component three since the highest number (0.6853 in Table **??**) is observed at component 3, and hence pH is given a weight of 0.225 = 0.16/0.71

Table 30: Principal Components (Eigenvectors)

	Comp1	Comp2	Comp3	Comp4
pH score	-0.25	-0.24	0.69	-0.27
EC score	0.37	-0.50	-0.084	0.33
C score	0.079	0.34	0.33	0.83
N score	0.018	0.65	0.31	-0.17
P score	0.54	-0.15	0.32	-0.048
K score	0.57	0.013	0.28	-0.25
S score	0.42	0.38	-0.38	-0.20

Note: Observations: 1003. Retained components: 4.

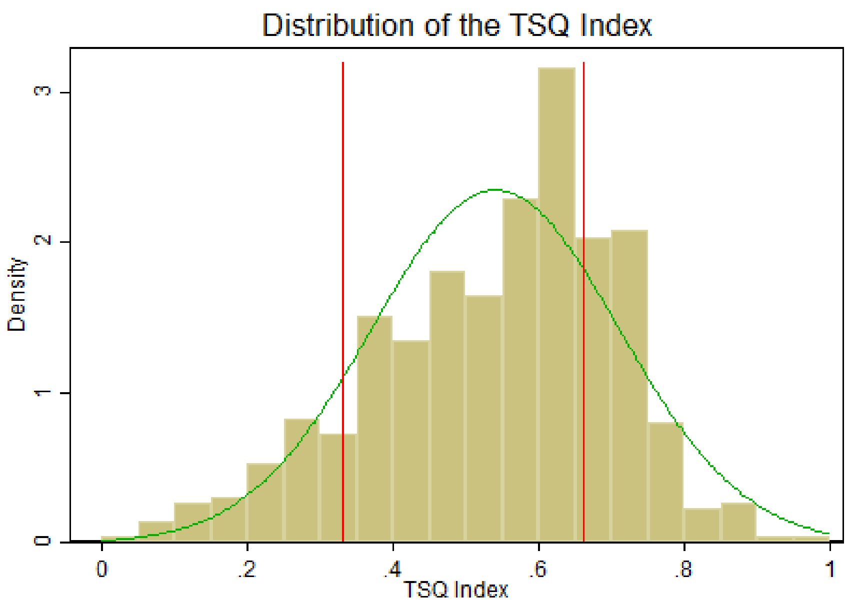

Fig 10: Distribution of the TSQ Index
Note: The vertical lines are the cutoffs based on which soils are classified as being of "poor", “fair”, or “good” quality.