Digital Information Provision and Behavior Change: Lessons from Six Experiments in East Africa*

Raissa Fabregas, Michael Kremer, Matthew Lowes, Robert On, Giulia Zane†

February 5, 2022

Abstract

Mobile phone-based informational programs are popular worldwide, though there is little consensus on how effective they are at changing behavior. We present causal evidence on the effects of six mobile-based agricultural information programs implemented in Kenya and Rwanda. The programs shared similar objectives but were implemented by three different organizations and varied in content, design, and target population. With administrative outcome data for over 156,000 people across all experiments, we are sufficiently powered to detect small effects in real input purchase choices. Combining the results of all experiments through a meta-analysis, we find that the odds ratio for following the text-based recommendations is 1.20 (95% CI: 1.14, 1.26). We cannot reject similar effects across experiments and for different agricultural technologies. We do not find evidence of message fatigue or crowd-out of other inputs. The effects, however, seem to diminish over time. Providing more granular information, supplementing the texts with in-person calls, or varying the messages' framing did not significantly increase impacts, but message repetition had a modest positive effect. While the overall effect sizes are small, the low cost of text messages can make these programs cost-effective.

*Funding from ATAI, the Dioraphte Foundation, Unorthodox Philanthropy, 3ie, and an anonymous donor made this study possible. We thank our partner organizations: the Kenya Agriculture and Livestock Research Organization (KALRO), and in particular to Dr. Martins Odendo, One Acre Fund (OAF), Precision Agriculture for Development (PAD), and Innovations for Poverty Action - Kenya (IPA-K) for all of their support during implementation and data collection. We are also grateful to OAF, the Mumias Sugar Company, and German Agro Action (Welthungerhilfe) for sharing soil data and specially to Dr. Javier Castellanos for providing the agronomic recommendations to PAD and Dr. David Guerena for the agronomic recommendations for OAF. We thank Lillian Alexander, Carolina Corral, Tomoko Harigaya, Charles Misiati, Chrispinus Musungu, Cara Myers, Alex Nawar, Carol Nekesa, William Wanjala Oduor, Violet Omenyo, Victor Perez, and Megan Sheahan for excellent project and research support. This project benefited from feedback from Rema Hanna, Rohini Pande, Kashi Kafle, Jessica Zhu, Jack Marshall, Andrew Dustan and participants at AFE, AAEA, MWIDC, the Texas Experimental Association Symposium, NEUDC, Sao Paulo School of Economics, Development Inovation Lab, Y-RISE and JILAAE. Disclosure: Michael Kremer is a board member of PAD. Giulia Zane's post-doctoral fellowship was funded by PAD. Fabregas, Kremer, and On did not receive any financial compensation from PAD but the organization partly funded this research. On worked as a volunteer for OAF at the time of the experiment. Lowes worked for OAF. Experiment registration ID: AEARCTR-0006837.

†Fabregas: The University of Texas at Austin, LBJ School of Public Affairs, rfabregas@utexas.edu; Kremer: The University of Chicago, Economics Department, Harris School and NBER, kremermr@uchicago.edu; Lowes: One Acre Fund, matt.lowes@oneacrefund.org; On: Independent scholar, roberton@gmail.com; Zane: World Bank, gzane@worldbank.org.

1 Introduction

The rapid diffusion of mobile phones in developing countries over the past few decades has unlocked new opportunities for governments and other organizations to disseminate information at scale in pursuit of their policy objectives. As a result, hundreds of digital initiatives have been deployed to address informational or behavioral barriers and change individual behavior (GSMA, 2020). While only a fraction of these initiatives have been evaluated, there is a growing literature assessing the effectiveness of these programs across a range of sectors, from health (Hall et al., 2014; Jamison et al., 2013), education (Aker et al., 2012; Cunha et al., 2017; Angrist et al., 2020) and finance (Karlan et al., 2012, 2016) to governance (Dustan et al., 2018; Buntaine et al., 2018; Grossman et al., 2020) and agriculture (Aker et al., 2016; Fafchamps and Minten, 2012; Cole and Fernando, 2021).

Much of the empirical evidence on the impacts of these programs on recipient behavior has been characterized as mixed (Aker et al., 2016; Deichmann et al., 2016; Baumüller, 2018; Grossman et al., 2020; Steinhardt et al., 2019). If intervention effectiveness is very sensitive to the specific features of a program, the identity of the implementing organization or recipients, or the local context, it might be difficult to draw broader policy conclusions about whether to scale up or extend an intervention to a new setting (Pritchett and Sandefur, 2015). However, the perceived mixed results could also stem from other methodological issues, such as sampling variation (Meager, 2019), selection biases (Glewwe et al., 2004), varying levels of statistical power (Ioannidis et al., 2017), differences in instruments, or publication biases (DellaVigna and Linos, 2020).

This paper examines the impact of informational digital interventions on behavior change by presenting new experimental evidence about the impact of six text-message-based agricultural extension programs on individuals' decision to acquire recommended inputs. Text messages are inexpensive and can reach basic phones without internet connectivity, making them a particularly attractive option for delivering information in low-income countries where smartphones are not yet widely used.[1] Despite this potential, texting might be too impersonal, light-touch, or restrictive to meaningfully convey information. Illiteracy, mistrust, or mistar-

[1]In 2020, we documented that some services in Kenya charged less than $0.006 per text message. In India, it varied from $0.006 to $0.0004, depending on the number of messages bought. From the point of view of carriers, the marginal costs of a text message are close to zero.

geting might also limit the effectiveness of these types of programs (Aker, 2017), especially at scale (Bird et al., 2019).[2]

The programs studied were implemented in Kenya and Rwanda by three different organizations: a public agency, a social enterprise, and a research-oriented non-profit. All the programs aimed to increase farmer experimentation with locally recommended agricultural inputs by conveying information through texts. Despite sharing similar objectives, the programs varied in other dimensions, such as user recruitment strategies, message content and design, implementation seasons, and complimentary access to in-person support. This set-up allows us to experimentally estimate impacts for each program individually and aggregate the results through a meta-analysis. The meta-analysis allows us to increase statistical power and test for impact heterogeneity across studies. This study set-up also captures a common occurrence in program implementation: organizations with similar tools and objectives will respond to their specific constraints by designing their programs differently. When considering scalability, it is crucial to understand to what extent these implementations details are critical for effectiveness.

Two features of this study are worth highlighting. First, we present evidence of programs with substantial sample sizes. In total, over 156,000 individuals participated across all six experiments. Results from low-powered studies can be mistakenly interpreted as evidence of no effects if they fail to detect small impacts (Ioannidis et al., 2017; Dahal and Fiala, 2020; McKenzie and Woodruff, 2014). This issue is particularly problematic for very cheap interventions, such as text messages, since the effect sizes required for these programs to be cost-effective are generally very small. Second, across all experiments, we use actual input purchases as

[2]There is growing experimental evidence on the impacts of text-message-based programs arising from a variety of sectors. By far the most empirical evidence comes from evaluations of health programs. Some examples include a program for adolescent reproductive health messages that found significant effects on knowledge, but no significant effects on reported intercourse or pregnancy (Rokicki et al., 2017) and positive impacts of SMS reminders on adherence to antiretroviral treatments (Pop-Eleches et al., 2011). Broader meta-analyses of global text-message health interventions suggest positive effects on health behaviors (Hall et al., 2015), though the evidence from low and middle-income countries remains limited (Hall et al., 2014). For financial behaviors in developing countries, text reminders for microloan repayment were reported to be insignificant unless the message included the loan officer's name (Karlan et al., 2015). In a different study, bank clients assigned to receive monthly saving reminders were three percentage points more likely to meet their commitment (Karlan et al., 2016). The effects of text messages with moral content reduced credit card delinquency by four percentage points, but researchers did not find other content to be statistically significant (Bursztyn et al., 2019). In governance, text messages increased bureaucrat policy compliance by four percentage points (Dustan et al., 2018), and Aker et al. (2017) showed that an SMS campaign could increase voter turnout. In education, text messages about student truancy increased school attendance by two percentage points (Cunha et al., 2017). In Peru, Chong et al. (2015) report no statistically significant effects of text messages on recycling behavior.

our primary measure of behavior change. We use administrative data to observe input purchases for farmers associated with the social enterprise that directly sells them inputs. For independent farmers, unattached to any organization, we provided discount coupons and partnered with dozens of small agricultural shops in the region that kept track of coupon redemption. Using purchase data mitigates the risk of social desirability or courtesy bias in self-reports, a particular concern for informational interventions (Baumüller, 2018; Haaland et al., 2020).[3] Using survey endline data for four programs, we can also compare self-reports with the administrative records and investigate other outcomes such as knowledge increases and potential crowd out in the use of non-recommended inputs.

Combining the effects of all six programs in a meta-analysis, we find that the odds ratio for following program recommendations is 1.20 (95% CI 1.14 to 1.26, N=6). The aggregate effect for following recommendations about a newly introduced technology (agricultural lime) is 1.22 (95% CI 1.13 to 1.31, N=6), and the effect for following recommendation on a relatively known technology (chemical fertilizers) is 1.32 (95% CI: 1.19 to 1.47, N=4). With only six studies, we cannot be conclusive about the extent of program heterogeneity. However, while we observe that some individual experiments had statistically significant impacts and others did not, we cannot reject the hypothesis that the effects were the same across programs and that the estimated effects differed mostly because of sampling variation. We estimate that, for our main outcomes, over 70% of the observed variation in treatment effects across studies is driven by sampling variation. We do not find evidence to suggest that the recommended input purchases crowded out the use of other categories of non-recommended inputs, though we reject the null hypothesis of homogeneous effects.

While the programs were designed as a way to provide new information to farmers and not just to serve as reminders or nudges, the evidence suggests that the effects operate partly through behavioral channels. For the programs for which we have survey data, we find positive effects on knowledge measures. Treated farmers were significantly more likely to correctly identify the purpose of the newly introduced input. However, we find that the effects on input purchases decayed once the program ended, although re-treating individuals with the same messages helped sustain impacts. This finding is in line with studies in other contexts that have found that well-timed reminder messages can be effective at altering behavior (Karlan

[3]Practically, using administrative data also helps with the financial costs of collecting individual surveys with very large sample sizes.

et al., 2015, 2016; Raifman et al., 2014). Moreover, we cannot reject that the impacts were the same for farmers regardless of their baseline knowledge about these technologies.

Adding to the literature on behavior measurement (Chuang et al., 2020; Karlan and Zinman, 2012) and the work that has found discrepancies between self-reported and actual behavior (Friedman et al., 2015; Karlan and Zinman, 2008), we find that the effects on the use of the newly introduced input were significantly larger when estimated using self-reported data relative to administrative purchase data. While we cannot say conclusively that these differences are due to misreporting, a substantial fraction of farmers who appear to over-report usage indicated that they had acquired inputs from sellers who, according to our records, had not stocked that input during the same period. This result signals the risk in relying on self-reported data to assess behavior change, even when the reported behavior is not particularly sensitive. Further, it is difficult to predict the direction of these inconsistencies: the discrepancy affects some but not all programs, and we find no difference in reports for other well-known inputs.

Finally, we use individual project experimental variation to draw additional lessons about the importance of different programmatic features and estimate potential spillovers to other farmers. Overall, we do not find strong evidence that message framing or the intensity of communication makes a critical difference in following the recommendations. First, we cannot reject that messages crafted using behavioral insights (e.g., loss/gain framing, sense of urgency, self-efficacy, social comparisons, etc.) were as effective as a standard message. We also do not detect additional gains from sending messages with more detailed information, such as highlighting that the recommendations were based on local soil data.

Second, to test whether in-person communication could help farmers make sense of the new information and significantly strengthen effects, one experiment complemented the text messages by randomizing a phone call from an extension officer. We find no evidence of additional significant impacts from this add-on. Message repetition, however, was modestly effective at increasing purchases.[4] Third, we estimate spillovers for the programs that targeted users who were part of farmer groups. We find effects of up to a third of the magnitude of the direct effects from treatment, suggesting that ignoring these externalities is likely to

[4]The sample sizes used on the experiments on message repetition were much larger than those in the in-person call, and better powered to detect small impacts. However, given the costs of in-person calling are unlikely to make this approach cost-effective, given potential effect magnitudes.

underestimate the benefit of text-message-based campaigns.

Without data on yields, we are unable to draw firm conclusions on the effects of these programs on farmers' profits.[5] However, using agronomic estimates of the impacts of these recommended inputs on maize yields in the region, we can provide a back-of-the-envelope calculation of the potential direct benefits of sending these text messages. Our estimates suggest that the benefit-cost ratio of sending these messages at scale is about 55 to 1.

This paper adds to the recent literature that finds modest but positive effects of low-touch interventions on behavior change (Benartzi et al., 2017; Oreopoulos and Petronijevic, 2019; DellaVigna and Linos, 2020).[6] In our case, however, we focus on digital informational interventions in low-income contexts. The effects of text messages compare favorably relative to more intensive, but also more expensive, programs such as in-person farmer events. Our results also speak to the literature concerned with the use of experimental evidence for policy scale up, particularly in development projects, where heterogeneity in treatment effects has been used as a measure of external validity (Pritchett and Sandefur, 2015; Allcott, 2015; Meager, 2019; Vivalt, 2016). We cannot conclude that there is heterogeneity in true impacts, and the results suggest that we should be cautious in qualitatively interpreting differences in significance across studies in the literature, because these differences could be driven simply by sampling variation and studies that are underpowered to detect small effects.

Finally, we complement expert qualitative summaries of the literature on digital interventions for development (Aker, 2017; Aker et al., 2016) and the handful of experimental studies assessing the impacts of digital agricultural extension systems. Larochelle et al. (2019) study a text-based program for potato farmers in Ecuador and find that the program increased knowledge and self-reported adoption of integrated soil management practices. Text messages sent by an agribusiness to sugar cane farmers in Kenya had positive yield impacts in one trial, but not in a second trial (Casaburi et al., 2014). Fafchamps and Minten (2012) report null

[5] Measuring impact on downstream outcomes, such as yields and profits, can be complex if the effect sizes that would make these types of programs cost-effective are small. Self-reported yields are noisy (Lobell et al., 2018) and objective measures such as physically harvesting a section of a farmer's plot could be prohibitively expensive to gather at the required sample sizes. The stochastic nature of rainfall and other features can further complicate this (Rosenzweig and Udry, 2020).

[6] An exception in this literature is work by Linos et al. (2020), who find precise null effects from six large-scale nudge experiments aimed at increasing tax credit take up in the U.S.

effects from a text-based program with weather, price, and advisory content in India.[7] A more sophisticated voice-based service, targeted at cotton farmers in India, increased the use of recommended seeds but no other inputs (Cole and Fernando, 2021). This paper expands what is known empirically about text-based agricultural extension programs, addressing some methodological limitations in the existing work.[8]

This paper is organized as follows: Section 2 presents the context and design of each program and their evaluations. Section 3 discusses the empirical strategy. Section 4 provides the main results, and section 5 discusses some of the additional lessons that we can draw from individual experiments. We present cost-effectiveness estimates in section 6 and conclude in section 7.

2 Context, Programs and Experimental Design

2.1 Context

The programs targeted maize farmers in Rwanda and Kenya between 2014 and 2017 (see Figure 1 for a map). In both countries, maize is farmed twice a year.[9] Maize is a staple food and traded commodity, and increasing smallholder productivity is an important policy objective to improve food security and reduce poverty. However, smallholder yields have remained low, partly due to soil degradation, soil acidity, and the low adoption of productivity-enhancing technologies (FAO, 2015).

High soil acidity, corresponding to pH levels below 5.5, can dramatically reduce crop yields by limiting nutrient availability to the plants (The et al., 2006; Tisdale et al., 1990; Brady

[7]As a comparison, we highlight relative differences in sample sizes with these other evaluations of text based-interventions. For instance, in Larochelle et al. (2019), n=353; Casaburi et al. (2014) n=1,849 and 2,819; and in Fafchamps and Minten (2012) n=1,000.

[8]A couple of additional literatures are worth highlighting here. First, the literature on the broader effects of mobile phone access on market performance and productivity (Jensen, 2007; Gupta et al., 2020; Aker and Mbiti, 2010; Aker and Fafchamps, 2015). Second, another set of papers has studied the market effects of providing information about crop prices through mobile phones (Camacho and Conover, 2010; Mitra et al., 2017; Nakasone et al., 2014; Courtois and Subervie, 2014; Svensson and Yanagizawa, 2009).Finally, other digital extension approaches, like delivering information via video, tablets or smartphone apps, have also been shown to have positive effects on farmers' beliefs and behaviors (Tjernström et al., 2019; Van Campenhout et al., 2020; Arouna et al., 2019). However, until smartphone penetration increases, such approaches will likely require an in-person component or third party to deliver them.

[9]In Kenya, the primary agricultural season, the long rain season, runs from March until August, and a secondary agricultural season, the short rains season, from September to December. In Rwanda, the main season occurs from September to January, and the secondary season from March to August.

and Weil, 2004). The application of agricultural lime to the soil is one of the cheapest and most widely recommended methods to increase soil pH. Several public agencies and NGOs in Africa have advocated for the use of lime, and experimental plots conducted in Kenya suggest that lime application can increase maize yields by 5-75% (Kisinyo et al., 2015; Gudu et al., 2005; OAF, 2014).[10] Yet agricultural lime is not a widely known or used input. In Kenya, only 6% to 12% of farmers in our samples reported having ever used it at baseline, and in Rwanda only 6% had previously purchased it.

Chemical fertilizers are more widely used, but most farmers in the sample areas have used only a specific type of phosphate-based fertilizer, diammonium phosphate (DAP), and fewer farmers regularly experiment with other options such as top-dressing fertilizers, like calcium ammonium nitrate (CAN) and urea.[11] Experimental plots suggest that fertilizers, particularly top-dressing ones, can be profitable (Duflo et al., 2008; Kelly and Murekezi, 2000), and current national and international recommendations have shifted towards encouraging farmers to use fertilizers that best fit the local conditions (KSHC, 2014; NAAIAP, 2014). Therefore, several organizations have attempted to provide information to farmers about different fertilizer options to encourage experimentation.

2.2 Partner Organizations, Programs and Randomization

In this section, we briefly summarize the characteristics of the implementing organizations, their programs and the main features of each evaluation.[12] The common treatment across all programs was information provision about agricultural lime. Four programs also sent information about locally recommended chemical fertilizers. Table 1 summarizes the six programs, and Table 2 briefly describes the characteristics of each experiment. A full description of each program and additional details of each evaluation can be found in Appendix H.

[10] These estimates reflect results for trials with and without combining lime with other inputs, particularly fertilizers containing nitrogen and phosphorous. For instance, OAF's experimental plots suggest that broadcasting lime evenly over maize fields before the planting season begins in combination with their standard fertilizer package increased yields by 25%.

[11] Top-dress fertilizers are applied to maize once the plant has started to mature. The application is timed to provide the plant with a boost of nitrogen at a biologically beneficial time.

[12] We define an 'implementing organization' as the primary organization that was in charge of designing the programs and delivering the messages. Each implementing organization faced its own constraints, goals and directives. IPA and PAD-affiliated researchers were involved in the analysis and evaluation of all six programs.

Table 1: Program Characteristics

	KALRO (1)	IPA/PAD1-K (2)	IPA/PAD2-K (3)	OAF1-K (4)	OAF2-K (5)	OAF3-R (6)
Org. Type	Public	NGOs	NGOs	Social Enterprise	Social Enterprise	Social Enterprise
Location	Kakamega and Siaya (Kenya)	Busia and Kakamega (Kenya)	Busia, Bungoma, Kakamega & Siaya (Kenya)	Busia and Kakamega (Kenya)	Bungoma, Busia, Kakamega and Vihiga (Kenya)	Western, Eastern, Southern (Rwanda)
Agricultural Season	SR 2015	SR 2016/LR 2017	LR 2017	SR 2016/LR 2017	SR 2017/LR 2018	Main Season 2017/2018
Recruitment	Farmers drawn from village census	Former NGO and contract farming participants	Clients of agrodealers	OAF clients in LR 2016	OAF clients in LR 2017	OAF clients in 2017
Eligibility	Phone owner, farmed during past year, in charge of farming	Planted maize in 2016, reside in program area	Clients of agrodealers	OAF clients in LR 2016	OAF clients in LR 2017	OAF clients in 2017
Message Content	Lime, fertilizer, seeds, field management	Lime, fertilizer, field management	Lime and fertilizer	Lime	Lime and fertilizer	Lime
Number of Messages	20 total (2 acidity/lime; 5 fertilizer)	24-28 total (7-9 acidity/lime; 4-9 fertilizer)	13 total (6 acidity/lime; 4 fertilizer)	6 total (6 acidity/lime; 0 fertilizer)	1-10 total (1-5 acidity/lime; 1-5 fertilizer)	1-4 total (1-4 acidity/lime; 0 fertilizer)
Lime recommended?	All (if acidic)	0.81	0.76	All	All	All
Key Fertilizers recommended	DAP, NPK, CAN, Mavuno	Urea	Urea	-	CAN	-
Used Local Soil Data?	No	Yes	Yes	Yes	Yes	Yes
Additional Services?	No	No	Phone-call	OAF Services & Call-center	OAF Services & Call-center	OAF Services & Call-center
Any Message Repetition	No	Yes	Yes	Yes	Yes	Yes
Opt-in	1	0.95	0.95	-	-	-
Previous lime use[b]	0.06	0.12	0.09	-	-	0.06
Previous season fert. use (any/recommended)[b]	0.83/0.83	0.92/0.18	0.88/0.19	0.95/-	0.93/0.15	0.95/-
Female[b]	0.65	0.37	0.34	0.64	0.69	-
Primary School[b]	0.53	0.60	0.72	-	-	-

Notes: SR denotes Short Rain Season (August-January) and LR Long Rain Season (March-August) [b] denotes data for control group at baseline. - denotes that data is unavailable. Lime recommended indicates whether all farmers received messages recommending positive amounts of lime, or the fraction that did. Fertilizer recommended whether fertilizer messages were sent, and if yes, the types of fertilizer. Opt-in indicates the fraction of farmers who when invited agreed to received texts.

2.2.1 KALRO

The Kenya Agriculture and Livestock Research Organization (KALRO) is a Kenyan semi-autonomous public agency with the mandate to promote agricultural research and dissemination. Its text-message program was developed in partnership with the Kenyan Ministry of Agriculture and was envisioned as a low-cost way to reach farmers with simple messages to help them adopt locally-adapted inputs and practices. A total of 20 agriculture-related messages were sent during the 2015 short rain season, of which two related to lime and five to fertilizer use. KALRO designed messages that were agronomically correct but relatively technical.[13] This is in line with the observation that governments often provide agricultural information that might be too technical or insufficiently actionable when communicating with the average farmer (Fabregas et al., 2019).

Participating farmers were recruited by field agents who went door-to-door in KALRO's catchment areas. Among all identified farmers, 95% met the inclusion criteria for the study (i.e., phone owner, responsible for farming, and had planted maize during the previous season) and were invited to complete a baseline survey.[14] Farmers were then randomized at the individual level into a treatment or a control arm. All farmers invited to participate in the text message program opted in. The final sample consists of 834 farmers. About two thirds of this sample is female, and at baseline only 6% reported ever using lime but over 80% had used one of the recommended fertilizers during the previous season.

2.2.2 IPA and PAD

Innovations for Poverty Action (IPA) is a research and policy organization, and Precision Agriculture for Development (PAD) is a non-profit organization that supports the provision of phone-based information services to smallholder farmers in developing countries. We discuss the impacts of two programs (IPA/PAD1-K and IPA/PAD2-K) implemented in Kenya through a partnership between these two organizations.

An important objective for both programs was to make messages actionable for farmers.

[13] E.g. the message related to lime read: *'If soil is acidic (pH less than 5.5), apply the recommended rate of agricultural lime at least 30 days before planting'*. Few farmers have tested their own land, so it would be challenging to know their own soil pH.

[14] As part of this project, a second treatment arm, testing in-person farmer field days, was also evaluated. The results are described in Fabregas et al. (2017b).

The program sought to address two issues: knowledge of soil acidity and lime application rate. Few smallholder farmers conduct soil chemistry tests on their farms. Soil testing is relatively expensive and not easily accessible, making it difficult for farmers to ascertain their own pH levels and soil chemistry.[15] Therefore, area-level soil information was used to make predictions about local soil acidity and give farmers more specific guidance.[16] Appendix I describes the soil data used to construct the messages. Based on this, both IPA/PAD programs recommended lime to those in areas where the median pH was under 5.5 (corresponding to 81% in IPA/PAD1-K and 76% in IPA/PAD2-K). Second, agricultural lime is cheap, but it is bulky. Lime can be difficult for farmers to store and transport in quantities sufficient to broadcast. Therefore, farmers were advised to micro-dose lime instead.[17] In experimental plots, lime micro-dosing increased yields by 14% (OAF, 2014). This approach enables a lower annual dosage but requires re-application each season.

(i) IPA/PAD1-K. The first IPA/PAD text-message program was implemented during the 2016 short rain and 2017 long rain agricultural seasons. The experiment consisted of two treatment arms and a control arm. During the first season, one treatment arm (*General SMS*) received messages with general advice but did not refer to local soil data. The second arm (*Specific SMS*) received specific information based on area-level soil data and guidance on recommended input quantities. Between 24 and 28 messages were sent during the 2016 short rain season, of which 7 to 9 messages dealt with soil acidity and lime, 4 to 9 messages were about fertilizers, and the rest covered topics related to other management practices.

Participating farmers were identified using existing farmer databases.[18] A sample of 1,897 farmers completed a short baseline survey over the phone and were later randomized into either treatment groups or a control group. Among those randomized into the treatment groups, 95% agreed to receive the messages. During the following agricultural season, both treatment groups received five identical messages promoting the use of agricultural lime. Overall, the sample for this experiment was 37% female, and at baseline, only 18% had ever

[15] A wet soil test can cost between $11-$30.

[16] In a separate project, we document that using area-level means rather than global means reduced the mean squared error of the prediction by 12% for pH (Fabregas et al., 2017b). Farmers were also advised to always experiment in a small portion of their land.

[17] A micro-dosed application targets the input at the planting hole or base of the plant. This avoids using inputs on soil between plants or between rows.

[18] Farmers in Busia were invited through an existing IPA database for an existing large-scale farming project. Farmers in Kakamega were part of a database kept by a large agrobusiness in the area.

used one of the recommended fertilizers. Approximately 12% reported ever using lime.

(ii) IPA/PAD2-K. A second program was implemented in the long rain season 2017 with a different sample of farmers and in two new areas. This program sent 13 messages solely focused on lime and fertilizers. Farmer recruitment was done via agricultural supply dealers (agrodealers) who invited existing clients to register to participate.[19] All registered farmers were then contacted over the phone to obtain consent and complete a short baseline survey. The final sample consisted of 5,890 farmers.

The farmers were randomized into one of three treatment arms or a control group. The first arm received only text messages (*SMS only*). To investigate whether real-time communication with a field officer could strengthen the texts, both an arm where farmers received a phone call by an extension officer (*SMS + Call*) and an arm where farmers could request to receive a call (*SMS + Call Offer*) were tested. This sample was 34% female, 9% had used lime in the past, and 19% had used the key recommended fertilizer.

2.2.3 OAF

One Acre Fund (OAF) is a social enterprise operating across six countries in Eastern and Southern Africa. OAF's model relies on training farmers in modern agricultural techniques and providing them with inputs on credit early in the agricultural season, which they later repay. OAF clients form groups of eight to eleven farmers who participate in the program together and are supported by a local OAF field officer.

To address the problem of high soil acidity, OAF offers farmers agricultural lime as an optional input. However, demand for lime was relatively low across their operating locations. Hypothesizing that this low demand could reflect a lack of awareness, OAF implemented two text-message programs in Kenya and one in Rwanda to encourage lime use (OAF1-K, OAF2-K, and OAF3-R).[20] To build area-level acidity recommendations, OAF used its own soil data

[19]This method offered several advantages. First, it was a low-cost and quick method to recruit farmers. Second, farmers who are clients of agricultural supply dealers might already be more likely to acquire inputs, be less credit constrained, and therefore benefit from an information-based program.

[20]Relative to farmers in other samples who rarely had contact with extension officers, OAF farmers receive intensive agricultural extension training. One goal of using a digital approach, however, was to devise a cheap way to convey new information that did not require additional training and delivery by OAF field officers, who already followed detailed and lengthy training protocols.

(see Appendix I).[21]

(i) OAF1-K. OAF's first text-message program was implemented in western Kenya during the 2017 and 2018 long rain seasons. The program sent only 6 messages about lime. Farmers were randomized into either of two treatment arms or a control. The first arm sent simple text messages alerting the recipients about soil acidity and encouraging them to use lime (*Broad SMS*). A second group received more detailed messages that mentioned the predicted level of acidity in the area, the amount of lime recommended and the expected returns to its application (*Detailed SMS*). Participants were randomly selected from lists of previous OAF clients in Busia and Kakamega counties. A final sample of 4,884 farmers participated in the experiment. At the end of the first season, farmers were cross-randomized across treatment and control groups to receive another message encouraging lime adoption in the subsequent season.

(ii) OAF2-K. A second program was implemented in the Kenyan counties of Bungoma, Busia, Kakamega and Vihiga during the 2018 long rain season. A total of 32,572 farmers participated in this experiment, all of them recruited through listings of prior OAF clients. Farmers were randomized at the individual level into a comparison group or one of two treatment arms: a *Lime only* group, which received messages only concerning lime, and a *Lime + CAN* group, which received messages about lime and the top-dressing fertilizer CAN. The larger sample size made it possible to cross-randomize the message framing and message repetitions (1 to 5 messages) among the treated farmers.The different framing versions included a basic message and messages that highlighted yield increases, encouraged experimentation, made social comparisons and promoted self-efficacy. Moreover, the message content was cross-randomized to send messages addressing the whole family instead of the individual. During the second agricultural season, OAF sent messages to both control and treated farmers, effectively ending the experiment.

(iii) OAF3-R. The third OAF program was implemented across Rwanda in 2017 and 2018. This program consisted of sending only lime-related text messages. The sample included 110,400 farmers, all previous OAF clients, and was designed as a two-staged randomized experiment, which enabled the identification of the average spillover effects. First, farmer groups were randomized into a full control group (*Full Control*), where no farmer received

[21] In addition, all of their programs offered a hotline to treated farmers. Farmers could call if they had more questions about lime. Take-up of this hotline was extremely low, with less than 1% of farmers using this service.

messages; fully treated groups (*Full Treatment*), where all farmers (who owned phones) received messages and a partially treated group (*Partial Treatment*), where farmers were further randomized into receiving messages or remaining as controls. This design allows us to study the extent of spillovers by comparing the outcomes of untargeted farmers in partially treated groups against those of farmers in full control. To study the direct effects of the program, we exclude within-group controls in the partially treated group. Similar to OAF2-K, message framing and repetition was randomized among the treated farmers.The different framing versions included a simple general message and messages that mentioned yield impacts, encouraged self-diagnosis, used soil data, explained how lime works, encouraged farmers to order immediately, highlighted issues of soil acidity and impact, and emphasized yield changes. The messages were further cross-randomized to be framed as a loss or a gain in yields. The following year, treatment assignment was re-randomized across all farmer groups.

2.3 Data

2.3.1 Baseline Data

An in-person baseline survey data was collected for the KALRO sample before the randomization took place. A phone-based baseline survey was completed with farmers in the IPA/PAD1-K and IPA/PAD2-K programs. All these surveys asked about demographics, prior agricultural practices, and input use. For the OAF projects, we rely on client administrative data from the previous seasons, which reported gender and previous input purchases from OAF.

2.3.2 Endline Data

We have at least one administrative measure of input purchases for each program to measure effects on farmer behavior. For KALRO and IPA/PAD, we use data from coupons redeemed at local agrodealers. The coupons were devised as a way to collect information on real input choices while minimizing experimenter demand effects. For the OAF projects, our primary measure consists of farmers' agricultural input orders placed with the organization. Additionally, we collected endline survey data for KALRO, IPA/PAD1-K, IPA/PAD2-K, and OAF1-K. We provide additional details below.

KALRO. At the end of the 2015 agricultural short rain season, farmers in the KALRO sam-

Table 2: Research Design

	KALRO (1)	IPA/PAD1-K (2)	IPA/PAD2-K (3)	OAF1-K (4)	OAF2-K (5)	OAF3-K (6)
Unit of randomization	Individual	Individual	Individual	Individual	Individual	Cluster (farmer group)
Sample Size	834	1,897	5,890	4,884	32,572	110,400
Treatment Arms (#)	1	2	3	2	2+	2+
Treatment Arms	**1.SMS**	**1.General SMS** and **2.Specific SMS**: sent additional information about local acidity level, input prices and quantities.	**1.SMS only**, **2.SMS + Call**: also received call by field officer, **3.SMS + Call offer**: offer to receive phone call	**1.Broad SMS**, **2.Detailed SMS**: additional info on degree of soil acidity, lime quantity, cost, and predicted yield increase.	**1.Lime only**, **2.Lime + CAN**: additional messages encouraging to buy extra CAN. **Cross-randomized**: message framing, repetitions, and frequency.	**1.Full treatment**: all farmers in a group got SMS. **2.Partial treatment**: half farmers in group got SMS. **Cross-randomized**: message framing, repetitions, and frequency.
Second Season SMS	No	Yes, maintain treatment status	No	Yes, cross-randomized	Yes, all	Yes, cross-randomized
Admin Outcome	Coupon (paper), LR 2016	Coupon (digital), SR 2016/LR 2017	Coupon (digital), LR 2017	OAF admin, LR2017/LR2018	OAF admin, LR2018/LR2019	OAF admin, 2017/2018
Coupon Value	50% discount lime, 50% discount fertilizer	Choice 10 Kg lime or soap (first season); 15% discount lime (second season); 30% discount CAN or urea	15% discount lime; 15% discount fertilizer	-	-	-
Baseline Survey	Yes	Yes (phone)	Yes (phone)	No	No	No
Endline Survey	Yes, SR 2015	Yes (phone), LR 2017	Yes (phone), LR 2017/SR 2017	Yes (phone), LR 2017	No	No

Note: All experiments included a control group in addition to the treatment arms. SR and LR denote the Short and the Long Rain agricultural season in Kenya, respectively. Treatment arms (#) denotes the number of treatment arms, for OAF '+' indicates that there were cross-randomizations in these samples for the number of messages (1-5), frequency sent, and framing (7 possibilities).

ple were visited and asked to complete an in-person endline survey. The survey contained knowledge and input use modules. During this visit, all treatment and control group farmers received two paper coupons redeemable for inputs at a discount at selected agricultural supply dealers in their nearest market center. The first discount coupon was redeemable for a 50% discount for agricultural lime. The second coupon was redeemable for a 50% discount for any chemical fertilizer of their choice (CAN, DAP, NPK, or Mavuno). All coupons had a unique ID that allowed us to link redemption to respondents. Participating agrodealers were asked to keep the coupons and record farmers' input choices and purchases. For the KALRO sample, the survey measures behavior changes concurrent with the program implementation, whereas the coupons measure input purchases for the subsequent agricultural season.

IPA/PAD. The IPA/PAD-managed programs sent input discount coupons via text message to all treatment and control farmers in their respective samples early in the season. Farmers could redeem the coupons with agrodealers in their preferred market center (reported at baseline). For the IPA/PAD1-K program, both lime and fertilizer coupons were sent during the 2016 short rains. The lime coupon was redeemable for either 10 kg of lime or 1 bar of soap. We intended to capture farmers' input choices without liquidity constraints by allowing farmers to choose between lime and another common product of similar value. The second coupon provided a 30% discount on top-dressing fertilizers. In addition, to measure effects over a second season, all farmers received lime coupons for a 15% discount in the 2017 long rain season. For this program, 32 agrodealers in 25 market centers participated in coupon redemption. A phone endline survey was conducted during the 2017 long rain season with this sample. The survey included questions about input use during the 2016 short rains and 2017 long rains.

The sample of farmers participating in the IPA/PAD2-K program also received two electronic coupons redeemable for a 15% discount for both lime and the recommended top-dressing fertilizer. The coupons were sent during the long rain 2017 season. For this program, we partnered with 102 agrodealers in 46 market centers. In addition, farmers in this sample were invited to complete a phone-based endline survey between the end of the 2017 long rains and the beginning of the 2017 short rains.

OAF. Outcomes for all OAF programs were measured through the agricultural input orders placed with the organization. One consideration is that the text-message interventions occurred before farmers could enroll with OAF for that particular agricultural season. Across all OAF interventions, between 60 and 76% of farmers who received text messages later enrolled to acquire inputs from OAF. While we do not find any evidence of differential OAF enrolling by treatment status (Appendix Table B7, panels D-F, columns (5)-(6)), we take a conservative approach and define our outcome variable as lime purchased through OAF, without conditioning on whether farmers were OAF clients at the time of the experiment. In addition, a random one-third of farmers in the OAF1-K experiment completed an endline phone survey during the 2017 long rain season.

3 Empirical Strategy

3.1 Validity of the Experimental Designs

Appendix Tables B1-B6 show baseline characteristics by treatment status together with tests of equality of means across treatment arms for each program. As expected, the treatment and control arms are balanced along most characteristics. We fail to reject the null that coefficients are jointly zero for all experiments, except for OAF1-K, where we find small differences at baseline: those in the treatment arms were more likely to purchase onion seeds, one percentage point more likely to purchase additional CAN fertilizer, and more likely to have received a repayment incentive the previous year. We control for all these variables in our main specifications, but the results are robust to their exclusion.

For each program, we also regressed an endline survey attrition dummy on treatment indicators. We do not find any evidence of differential attrition by treatment status (Appendix Table B7, panels A-D). The survey completion rates for the four programs that collected this type of data ranged from 79% (IPA/PAD1-K) to 92% (KALRO).[22]

3.2 Individual Program Impacts

Our primary outcomes are 'following lime' and 'following fertilizer' recommendations.[23] For programs for which we have access to survey data, we can also measure changes in agricultural knowledge and use of other inputs. In all cases, we estimate intention-to-treat (ITT)

[22]Note that only a random one third of farmers in the OAF1-K experiment were attempted to be interviewed.

[23]The variable 'following lime' is coded as one if the farmer used lime and lime was recommended or if the farmer did not use lime and lime was not recommended. The variable 'following fertilizer' is coded as one if the farmer used one of the fertilizers listed in Table 1.

effects.[24] The general equation we estimate for each program is:

$$y_i = \alpha + \beta Treatment_i + X_i \nu + \gamma_w + \epsilon_i, \quad (1)$$

where y_i is the outcome measure for farmer i. $Treatment_i$ denotes a dummy variable(s) indicating treatment, X_i is a vector of controls for farmer-specific characteristics, γ_w controls for area fixed effects and ϵ_i is the error term. The coefficient β estimates differences between treatment and control farmers. Since many experiments tested several treatment arms and message variants, our main results show estimates pooling all treatment arms together to increase power and simplify the analysis and discussion. However, we provide tables with results for each treatment arm in Appendix E and highlight some lessons from these experimental variations in Section 5.

For binary outcomes, we estimate a non-linear analogue to equation 1 using a logistic regression model and report the coefficient β in terms of odds ratios (OR) for the probability of acquiring the input.[25] In the appendix, we also show results for linear probability (LPM) specifications expressing effect sizes in percentage points.

To improve precision and address small baseline imbalances, we control for the strata used in each randomization, demographic characteristics, farming practices, previous input use, and location fixed effects, and for the survey data we include enumerator fixed effects. Appendix A contains a list of controls used in all regression specifications. In the OAF3-R experiment, the errors terms are clustered at the farmer group level.

3.3 Meta-analysis

To synthesize the evidence across these various experiments and present a weighted average of the study estimates, we combine the results in a meta-analysis. We use a random effects model, which assumes that true effects in each study are normally distributed. The weighted average effect, therefore, represents the mean of the distribution of true effects. Formally, the

[24] Some farmers might not have received the messages. For instance, network issues, incorrect phone numbers and uncharged phones might have limited reach. Unfortunately, we don't have administrative information on whether individuals opened the messages or not. This approach underestimates the impacts of receiving the messages.

[25] The Cochrane Handbook Chapter 6 (Higgins et al., 2019) discusses criteria for choosing effect measure units in a meta-analysis. For instance, it is preferred to have a summary statistic that gives values that are similar for all studies. In this sense, relative effect measures are, on average, more consistent than absolute measures. The more consistent the effects, the more reasonable it is to express the effect as a single number.

model can be written as:

$$T_j = \mu + e_j + \zeta_j \quad (2)$$

where T_j is the observed effect for study j, μ is the underlying true average effect, e_j represents the measurement error due to sampling variation and ζ_j is the difference between the average effect and the effect of program j. Moreover, $e_j \sim N(0, \sigma_j)$ and $\zeta_j \sim N(0, \tau^2)$. σ_j is the within-study standard error of the treatment effect estimate that is observed for each study, while τ^2 is the between-study variance in true effects that has to be estimated from the data. The estimate of μ is:

$$\hat{\mu} = \frac{\sum_{j=1}^{s} w_j T_j}{\sum_{j=1}^{s} w_j}$$

where w_j are study-specific weights given by the inverse of the variance. In this case,

$$w_j = \frac{1}{(\hat{\tau}^2 + \hat{\sigma}_j^2)}$$

In practice, we estimate τ^2 using the DerSimonan and Laird method (DerSimonian and Laird, 1986). We also show results using a number of alternative estimation methods (Hartung-Knapp-Sidik-Jonkman and Empirical Bayes). In addition to τ^2, we report two other measures of heterogeneity across programs: Cochran's Q statistic to test the null hypothesis of homogeneous effects across studies and Higgin's and Thompson's I^2, the percentage of variability not explained by sampling error (Higgins et al., 2003; Higgins and Thompson, 2002).[26] We also estimate 95% prediction intervals.[27] For situations where there are multiple outcomes

[26] The Q statistic is a chi-square statistic with s (number of studies) minus 1 degrees of freedom and is calculated by:

$$Q = \sum_{j=1}^{s} w_j (T_j - \frac{\sum_{j=1}^{s} w_j T_j}{\sum_{j=1}^{s} w_j})^2$$

The null is that all treatments are equally effective. This test, however, has low power when the number of studies is small (Higgins et al., 2008). The percentage of variability, I^2, measures the share of variability not explained by sampling error and is given by:

$$I^2 = max\left\{0, \frac{Q - (df - 1)}{Q}\right\}$$

I^2 is less sensitive to the number of studies included, but it depends on their precision (Borenstein et al., 2017a). While there is subjectivity on interpreting the magnitudes, Higgins et al. (2003) provides the following rules of thumb: I^2=25% for low, I^2=50% for moderate, and I^2=75% for high heterogeneity.We report I^2 and a corresponding 95% confidence interval.

[27] Prediction intervals provide a predictive distribution of future effects in exchangeable settings, accounting for uncertainty in the effect and spread of a random effects distribution. It is estimated through the formula $t*\sqrt{(\sigma^2 + \tau^2)}$ where t denotes the critical value from a student's t distribution.

per study, we compute the mean of the outcomes for each study and account for within-trial correlations (Borenstein et al., 2017a).

We complement this analysis in two ways. First, we pool all datasets together and estimate a single model (as in equation 1) with strata controls and experiment dummies. Second, in Appendix G, we present the meta-analysis results using Bayesian hierarchical random-effects models (Rubin, 1981; Gelman et al., 2013).

4 Main Results

A summary of the main meta-analytic results are reported in Table 3. We discuss them in this section.

4.1 Impacts on Awareness and Knowledge

First, we ask whether the text-messages had any impacts on the awareness and knowledge of agricultural lime. We focus on lime because it is a relatively unknown type of input and because it was the main focus of all of the programs.[28] Figure 2 show that the treatment effects as an odds ratio for farmers having heard of lime (awareness) is 1.23, but the average effect is statistically insignificant (95% CI 0.96 to 1.57). However, there is substantial heterogeneity in this result. The p-value of the Q statistic is 0.03 and I^2=65.9% (95% CI 0% to 88%). In contrast, the text messages increased the proportion of farmers who knew that lime was used as a remedy for soil acidity (knowledge). Across projects this was recorded as free text, without prompting, and coded into categories by the data entry team. The odds ratio for knowing that lime can reduce soil acidity is 1.58 (95% CI 1.41 to 1.78). We cannot reject the null of homogeneous treatment effects on knowledge. The p-value of the Q statistic is 0.51 and I^2=0 (95% CI 0% to 85%). Overall, while farmers might have heard about this input regardless of treatment status, text messages were successful in conveying information about the purpose of a new technology. Estimates from linear probability models are shown in Appendix Table C1.

[28]No equivalent questions were asked about recommended chemical fertilizers during the endline survey across projects.

4.2 Impacts on Following Input Recommendations

We start with our preferred estimates using administrative purchase data. This includes data concurrent with the first implementation season for all programs except for KALRO's, where we use results based on coupon redemption for the subsequent agricultural season. We discuss effects from survey data in the following subsection.

Agricultural Lime. We examine effects on the main variable that all programs were hoping to affect: following lime recommendations.[29] Individual program effects range from a statistically insignificant 0.90 (95% CI 0.54, 1.52) for KALRO to 1.39 (95% CI 1.14 to 1.71) for OAF1-K (Figure 3).[30] The combined odds ratio for following the lime recommendation is 1.22 (95% CI 1.13 to 1.31).[31] However, we cannot reject the null of homogeneous treatment effects (p-value=0.21). The results are robust to alternative methods to calculate τ^2 (Appendix Table G1, Panels B-C). Similarly, the bayesian meta-analytic estimate for the effect of following lime recommendations is 1.27, and we estimate that 71% of observed heterogeneity is sampling variation (Appendix Table G2), though we note that the confidence intervals are quite wide. The prediction interval, which gives a more intuitive sense of the range of effects of where a future sample would lie (Borenstein et al., 2017b), ranges from 1.03 to 1.45.

Appendix Table C2 shows the results employing a linear probability model. The meta-analysis yields a combined effect of a 2 percentage point increase in the probability of following the recommendations (95% CI 0.01 to 0.03).[32]

Fertilizers. Next, we examine the impact of these programs on purchases of recommended chemical fertilizers. Only four programs (KALRO, IPA/PAD1-K, IPA/PAD2-K, and OAF2-K) recommended fertilizers in addition to lime. Except for KALRO, which aimed to increase the use of relatively well-known and widely used fertilizers, the programs encouraged farmers

[29]For the PAD/IPA programs, we code the recommendation as being followed if the farmer used lime and lime was recommended or if the farmer did not use lime and lime was not recommended. The OAF programs recommended positive amounts of lime to all farmers. KALRO recommended lime to farmers if their soil was acidic. Since the program took place in an acidic region, we assume purchasing lime is equivalent to following lime recommendations for this sample.

[30]Effects sizes for OAF programs are slightly larger if we restrict to the sample of farmers who re-enrolled as OAF clients (Appendix Table C2, columns 3 and 6).

[31]Pooled data from all experiments into a single regression show qualitatively similar conclusions (Appendix Table D1, Panel A, column 1 and 3).

[32]Using percentage points, suggests a higher degree of heterogeneity, and we reject the null of homogeneous treatment effects across programs (Appendix Table G1, Panel A). This last result is driven by the inclusion of the Rwanda project, which resulted in a precisely estimated one percentage point increase in lime use. Yet, we note that the predicted range of program effects is reasonably small.

to experiment with less-known fertilizers. Key recommended fertilizers are listed in Table 1. Based on administrative data availability, we define 'following fertilizer recommendations' if farmers purchased at least one of the key fertilizers recommended by each program. This captures a shift towards recommendations, though it does not necessarily indicate an increase in overall fertilizer use if farmers substitute between different types of fertilizers. The odds ratio increase in the likelihood of following the fertilizer recommendations is 1.32 (95% CI 1.19 to 1.47) (Figure Figure 4a). We fail to reject the null of homogeneous effects (Q statistic p-value 0.71, I^2=0% (CI 0% to 85%)). Looking at the results in percentage points, we find an overall two percentage point increase in recommended fertilizer purchases (Appendix Table G1, Panel A). The bayesian results suggest a similar magnitude, 1.33, though the confidence intervals are wider (95% CI 0.94 to 1.81).

We also estimate effects on overall fertilizer purchases for which we have administrative outcome data. The results are shown in Figure 4b. The combined effect is 1.19 (95% CI 0.95 to 1.49). The smaller coefficient reflects substitution between types of fertilizers.[33] Altogether, the results are in line with the stated objectives of these programs: chemical fertilizers are well-known inputs, and there is some evidence to suggest that messages shifted farmers towards recommended blends.

4.3 Differential effects of self-reported vs. administrative data

Are effects measured using self-reported vs. administrative data equivalent? To answer this question, we can compare the results of four projects for which we have both types of data (KALRO, IPA/PAD1-K, IPA/PAD2-K, OAF1-K).

Figure 5a shows the meta-analysis of the ratio between the OR coefficients obtained using survey data and those obtained using administrative records for lime.[34] The meta-analytic estimate is 1.18 (95% CI 1.07 to 1.29), which indicates that the effects estimated using survey data

[33]Using survey data, we can further look at the effects of using any of the mentioned fertilizers, especially planting fertilizers that are well known in the region (and therefore, were not the focus of the IPA/PAD programs). Appendix Figure G1 shows effects on overall fertilizer purchases for all mentioned fertilizers, using administrative data if it exists, or survey data if it does not. The likelihood of purchasing any type of fertilizer is 1.15 (95% CI 0.97 to 1.37).

[34]The ratio of odds ratios compare the change in effects between two groups. A ratio of odds ratios greater than 1 implies that the effect was greater when measured with survey data than with administrative data. Standard errors are calculated allowing for correlation between the two estimates (Borenstein et al., 2009). We assume the correlation between the coefficients obtained with survey and administrative data is equal to 0.69, which is the correlation between the corresponding outcome variables for lime.

are significantly higher than those using administrative records. Which programs drive this difference? An obvious suspect is the KALRO program, since the administrative and survey data correspond to two different seasons. However, the difference between both data sources is relatively small (and statistically insignificant for both measures, Appendix Table C2). Similarly, the survey data lines up reasonably well with the administrative reports for OAF1-K. However, for the IPA/PAD programs, the survey results are statistically larger than the ones estimated using data from coupon redemption. This discrepancy could indicate either of two things. One possibility is that the survey data is affected by social desirability or recall bias, and that true lime purchases are misreported in the questionnaire. This could be the case, for instance, if farmers felt compelled to report that they followed the recommendations even when they did not. A second possibility is that the coupon redemption underestimates true lime use, since farmers might have acquired inputs from other sources not captured by the administrative data.

We explore these possibilities for farmers in the IPA/PAD2-K sample, for which we have more information. First, we check whether those farmers who were more likely to have other sources of lime (because they also reported participating in OAF programs) are more likely to report using lime but not redeeming the coupon. We find that within this sample, participating in OAF programs (35% of the sample) is associated with a 4 percentage point increase in the likelihood of reporting using lime in the survey but not redeeming the coupon (from 8 to 12%). This could suggest that some farmers could have procured lime from alternative sources, like OAF.

We can also compare farmers' reports about which shops they acquired inputs from against other surveys completed with agricultural supply dealers about their stock. We find that 64% of farmers who reported using lime in the survey (but who did not redeem the lime coupon, according to our records) also reported that they had acquired lime from a shop that had not sold lime during the same period, according to our shop data. This hints at misreporting in the farmers' survey data. Overall, true effects on lime use are likely to be between these two bounds, but we take the more conservative administrative results as our preferred estimates.

When considering discrepancies in fertilizer use, we have data for three programs (KALRO, IPA/PAD1-K, and IPA/PAD2-K). The mean fertilizer use in the control group is significantly higher in the survey than in the administrative data for all programs (e.g. 81% vs, 41% for KALRO, 15% vs. 2% for IPA/PAD1-K and 16% vs. 2% for IPA/PAD2-K). This could suggest that farmers procured fertilizer from sources other than the participating agrodealers that redeemed coupons. Relative to the discrepancy for lime, the direction of the gap between survey and administrative data for fertilizer is negative, though the overall difference is not statistically significant (Figure 5b).

4.4 Combined Effects on All Recommended Inputs and Practices

Previous sections focused on the uptake of lime and selected fertilizers because adopting these inputs was a key program objective. Consequently, the evaluations measured those outcomes through actual purchases. However, KALRO and IPA/PAD's programs sent information about other management practices and inputs. In this section, we report overall effects considering all possible adoption outcomes. In all cases, we use administrative data if it is available and survey data otherwise. Appendix Table A1 reports the list of the inputs recommended and measured for each program.

To consider multiple outcomes, we follow two approaches. First, we extend the meta-analysis model described above to incorporate multiple treatment effect estimates within studies, accounting for the fact that effects might be potentially correlated within a study (Borenstein et al., 2009).[35] Figure 6a shows the corresponding forest plot. The estimated odds ratio is 1.20 (95% CI 1.14 to 1.26, N=6), and we fail to reject the null of homogeneous treatment effects (p-value= 0.61). The prediction interval ranges from 1.11 to 1.28.

The bayesian estimate is also 1.20, and under that model, we estimate that 75% of the observed heterogeneity is sampling variation (Appendix Table G2). A limitation of conducting a meta-analysis with only six studies is that the confidence intervals for I^2 are quite large (0% to 84%), meaning that it is difficult to say anything conclusive about heterogeneity. However, by the same token, there is also no evidence to suggest the existence of true impact heterogeneity across these programs. It would be incorrect to simply conclude that the effects of these digital programs are 'mixed' simply because some exhibited statistically significant impacts and

[35]For each program, we calculate the average effect size as the average of the outcome specific log-odds, and derive its standard errors by assuming 0.17 correlation across effect sizes, which is the correlation of lime and fertilizer outcomes in the pooled sample. We perform sensitivity analyses of such assumption, both for the results in this section and the next, and find that all results are robust to different assumptions on the correlation across outcomes, including 0 and 1.

others did not.

As a second strategy, we standardize treatment effects for each experiment following the construction of indices as per Kling et al. (2007). Combining these point estimates through a meta-analysis, we find that the overall effect of the programs, expressed in terms of standard deviations, is 0.06 (95% CI 0.03 to 0.08) (Table G1, row 8).

We conclude that text messages affect farmers' choices. While the effects are relatively modest, the magnitudes are not substantially different from those of much more intensive and costly extension approaches. To put these effect sizes into perspective, consider the effects of other agricultural interventions. Large in-person extension events in Western Kenya increased the purchase of agricultural lime by four percentage points (Fabregas et al., 2017b). In India, Cole and Fernando (2021) find that a much more sophisticated voice-based service increased the adoption of a key recommended cotton seed by 0.09 standard deviations, and BenYishay and Mobarak (2013) find increases of up to 6% in pit planting using in-person extension services. In Uganda, a video-based intervention increased the use of chemical fertilizers by five percentage points (Van Campenhout et al., 2020).

4.5 Effects on Non-Recommended Inputs

Does following the recommendations crowd out the purchase of other non-recommended inputs? On average, we do not find that the programs affected the use of inputs that were not recommended. The list of the inputs we consider is reported in Appendix Table A1. The combined effects on non-recommended inputs are negligible and statistically insignificant: 1.00 (95% CI 0.92 to 1.09, N=5) (Figure 6b and Table 3, row 6).[36] However, we reject the null of homogeneity across results (p-value=0.06, I^2=55%). This is driven by the IPA/PAD2-K program, where we find that farmers reduced their consumption of other inputs such as hybrid seeds and pesticides.

[36] To account for multiple outcomes per program, we take the two approaches described in the previous section. As for all the recommended inputs, we assume that the correlation across inputs is 0.17, which is the correlation between lime and fertilizer in the pooled data. The results are robust to different assumptions, including 0 and 1. Conducting a meta-analysis that allows for within-study effect correlation and constructing indexes for each program lead to similar conclusions. Appendix table C6 shows the results by experiment, using the seemingly-unrelated regression framework to account for covariance across estimates. The results for the IPA/PAD1-K program suggest that farmers substituted away from other types of chemical fertilizers, in favor of those recommended by the program (Panel B, columns 5 and 6). The point estimate for other types of fertilizer is negative also for the IPA/PAD2-K program, although smaller and not statistically significant (Panel C, columns 5 and 6).

4.6 Effect Persistence

We measure input use during a subsequent agricultural season ($t + 1$) for farmers that were only treated during one season (t). We take this as a measure of effect persistence. Four programs allow us to study this question: KALRO, IPA/PAD2-K, OAF1-K, and OAF3-R. We use administrative data for all of them except for IPA/PAD2-K, for which we only have survey data for the second season. Coefficients are generally positive (except for OAF1-K) but statistically insignificant. Combining the results in a meta-analysis, we find that the effects are positive for both fertilizer and lime, but the magnitude is smaller than when effects are measured on the concurrent season. For lime the combined odds ratio is 1.08 (95% CI 0.98 to 1.19) and for fertilizer 1.09 (95% CI 0.99 to 1.20) (Figures 7a and 7b). We fail to reject the null of homogenous effects in both cases. While we cannot reject the null that these effects are equivalent to the ones measured in the first season, we take this as suggestive evidence of effect decay after the end of these interventions. Appendix Tables C4 and C5 show persistence results ('Treated S_t') using survey and administrative data separately for lime and fertilizer, respectively.

4.7 Message Fatigue

We also ask whether re-treating farmers during a second season sustained the program effects. We can answer this question by looking at lime use for the three programs that re-treated farmers during a second season: IPA/PAD1-K, OAF1-K, and OAF3-R. [37] We use administrative data in all cases. The combined effect for the three programs that re-treated farmers over a subsequent season is 1.33 (95% CI 1.19 to 1.49) (Figure 8) and we fail to reject the null of homogeneous treatment effects (p-value=0.75). The corresponding effect for those three programs measured over the first season is 1.26 (95% CI 1.15 to 1.38). These results provide little empirical support for the idea that re-treating farmers with text messages will lead to fatigue or message avoidance. Results for each program are listed in Appendix Table C4 ('Treated S_t & S_{t+1}').

[37] OAF2-K program is excluded because everyone in the control group was treated during the second season.

4.8 Who is most responsive to these programs?

A potential concern about digital-based approaches is that they will favor younger, richer or more educated farmers. However, we find little evidence of heterogeneous effects by gender, level of education, farm size or age (heterogeneity results for each program are shown in Appendix Tables F1-F2). For robustness and to increase power, we also show results pooling all data sets together in Appendix Table D2 and conduct meta-regressions (results not shown). We find no evidence of differential program effect by these characteristics. Moreover, we find that input purchase was not differentially affected by whether farmers had used or heard about fertilizer or lime in the past. We interpret this finding to suggest that some effects of text-messages operate through channels other than simply raising awareness about these inputs.

5 Lessons from Individual Experiments

This section uses individual projects' experimental variation to gather lessons about the effects of message content, repetition, complementary services, and spillovers.

5.1 Effects of Message Framing and Content

We use the experimental variation of four projects to draw out some lessons on framing and content. OAF2-K and OAF3-R randomized different versions of the input messages, with the intention of appealing to well-known behavioral biases or providing additional information (e.g., highlighting social comparisons, targeting self-efficacy, highlighting expected yields, nudging them to order immediately, etc.). Overall, we cannot reject the hypothesis that all messages were equally effective. Appendix Table E2 Panels A-B show effects for lime and fertilizer for OAF2-K, Panel C shows effects for lime for OAF3-R. Columns (1) and (5) present effects against each control group. Columns (2) and (6) show effects against the basic message. The only statistically significant effect we detect is for messages that included information on the potential increase in yields for OAF2-K, but the effect is only marginally significant and does not hold for fertilizer purchases.

All OAF2-K messages were further cross-randomized to address the whole family instead of the individual (e.g., the word "you" was replaced with "your family"). We find some evidence that addressing the entire family (*Family framed SMS*) was less effective for lime (Panel A, column (3) and (6)). The effects for fertilizer are also negative but not statistically significant (Panel B, column (3) and (6)). Finally, OAF3-R messages were cross-randomized to be framed either as a loss or a gain (e.g., "to increase yields" vs. "to avoid a yield loss"), but we do not detect any impact differences on lime purchases between these different phrasings (Panel C, column (3)).

A second piece of evidence of content variation comes from the IPA/PAD1-K and OAF1-K projects. The IPA/PAD1-K project randomized farmers to either a general (*General*) information arm or a treatment arm that provided additional information about the extent of soil acidity in the local area (*Specific*). While the *Specific* treatment arm was significantly more likely to increase knowledge about lime, we do not find significant differences between the arms on the probability of purchasing either input (Appendix Table E1, Panel A). These results are in line with the *broad* and *specific* treatment arms implemented by the OAF1-K program. The point estimates between treatment arms are similar, and we cannot reject equality (Appendix Table E1, Panel C).

Overall, providing additional details on local soil characteristics or specific message framing made little difference in whether farmers followed the recommendations. However, since the cost of optimizing messages is very low, this is an area that warrants further exploration.

5.2 The Effects of Message Repetition

OAF2-K and OAF3-R cross-randomized the number of repeated messages, where each message was sent every couple of days. We find evidence from both projects that repetition matters. One additional text increased the odds ratio of purchasing lime by 1.03 (Appendix Table E3, Panel A, column (4)) in the OAF2-K program and 1.07 in the OAF3-R program (Panel B, column (4)). In both programs, the effect is driven by receiving at least two SMS messages, and we find no effect from additional messages. The effect for an additional fertilizer message is 1.07 (Panel B, column (4)).

5.3 Are text messages strengthened by phone calls?

Text messages can be cheap and timely. Farmers might also consult them at later times if they forget content details. However, texts are a restrictive medium in which to convey information.

Receiving information through texts is also a fairly passive exercise. To address these concerns, the PAD/IPA2-K project experimented with three treatment arms: in the first arm, farmers received only text messages (*SMS*). Farmers received text messages and a phone call from an extension officer (*SMS+Call*) in the second arm. In the third arm, farmers received the text messages and were offered the possibility of texting back to receive a call (*SMS+Call Offer*). All calls were free.

We do not find high demand for an additional phone call. Only 15% of farmers assigned to the *SMS+Call Offer* group requested a call. This relatively low demand is in line with that of the OAF projects, where a hotline was also available for all treated farmers, but where less than 1% called the toll-free number. Moreover, while receiving a call was more effective in raising awareness about lime, we do not find statistically significant differences between any of the treatment arms in following lime recommendations (Appendix Table E1, Panel B). Overall, we do not find strong evidence to suggest that receiving a call made an appreciably large difference in behavior relative to sending text messages alone.

5.4 Are there information spillovers?

These programs could create spillovers if beneficiaries share information with non-participants who might also adopt the recommended technologies. If non-study farmers benefit from the text-message programs, we risk underestimating the overall impacts of these programs through these indirect channels. Moreover, since five projects relied on individual randomization, if farmers in the control group benefited from this information, we could also underestimate the direct effects of the programs.

To assess the potential magnitude of these spillover effects, we pursue three approaches. First, we leverage the OAF3-R randomization, which was designed to capture any potential within-farmer group spillovers.[38] Comparing untargeted farmers in partly treated groups against those in pure control groups, we find that the odds of purchasing lime increased by 15%. This effect is statistically significant and corresponds to a 1.04 increase in the odds ratio (or a 0.4 percentage point increase in the linear probability model) relative to the pure control group (Table C8, panel C, columns (2) and (6)). These estimates are about half to one third of

[38] Phone ownership is much lower in Rwanda than in Kenya (only 53% of the farmers registered had a phone number in OAF's database), so one of the objectives of this program was to measure spillovers among farmers in the same group, including those who did not own a phone.

the effect of the direct treatment, and suggest that the effects we measure are likely to be an underestimate of the effects of the programs. However, since OAF farmers operate together in groups, it's likely that spillovers are higher in this experiment than in projects where subjects operate individually (IPA/PAD and KALRO).

As a second strategy, we also explore whether farmers without registered phones in the treated groups in OAF3-R were more likely to adopt inputs relative to farmers without registered phones in non-treated groups (phone ownership is almost universal in Kenya, so we can't use this approach for the other projects). Again, we find evidence of spillovers to these groups, with an 18% increase in the odds of following the recommendations among those without phones (Table C8, Columns (4) and (8)).

Third, to complement these exercises, we use the variation created in the number of treated farmers within a group for the other two OAF programs. We estimate regressions of the following form for control farmers:

$$y_i = \alpha + \beta_1 Treat\_Peers_i + \beta_3 Group\_size_i + X_i \nu + \gamma_w + \epsilon_i, \tag{3}$$

where we include the number of group members who are assigned to the treatment ($Treat\_Peers_i$) and control for group size. In this case, β_1 compares control group respondents who are exposed to a higher fraction of treated farmers. Using this specification, we find a small positive effect for OAF2-K (for both lime and fertilizer messages) but no corresponding effect for OAF1-K or OAF3-R (Table C8, columns (1) and (5)). Overall, while there is some variation, the cleanest evidence indicates the existence of indirect gains for these programs, at least for those in which farmers often interact.

6 Cost-Benefit Analysis

We present two types of calculations to give a sense of the returns to these programs. First, we compare the costs and effects on lime adoption for text-based programs relative to non-digital programs with similar goals. While this comparison is not sufficient to inform the overall decision about whether to invest in these programs, it is helpful for making comparisons with other extension approaches if we take the policy objective as fixed. Second, we conduct back-of-the-envelope calculations to provide an estimate of the benefit-cost ratio of an intervention of this kind implemented at scale. To establish benefits, we combine information from the effects on lime and fertilizer adoption with existing agronomic data that allows us to estimate the impact on yields and agricultural profits.[39]

In both cases, we consider the marginal costs of the text messages for cost estimations. This assumes that other potential fixed costs of running these programs would be incurred with or without the text-message component.[40] The cost of sending one text message is approximately $0.01, and these costs can be significantly lowered to $0.001 if the programs operate at scale with bulk texting. We consider a program that sends four messages, three of them specifically about lime. Therefore, we estimate the marginal cost per farmer to be $0.04 - $0.004 per season.

Cost-Effectiveness. We estimate that for the text-based programs, the cost of getting one farmer to experiment with lime ranged between $0.20-$2 USD depending on the scale of the program (see Table 4 for details). We compare these estimates to those of in-person extension approaches implemented by KALRO and OAF in the region.

An experimental evaluation of KALRO's Farmer Field Days (FFDs) in Western Kenya, which consisted of large in-person meetings with farmers where they could observe test plots, increased the use of agricultural lime by 4 percentage points (Fabregas et al., 2017a), and amounted to a per-farmer cost of at least $9.[41] Given that the FFDs covered various topics, not only lime, we conservatively attribute 1/5 of their cost to the lime program and estimate that the cost per adopting farmer was approximately $45.

A second experiment conducted by One Acre Fund in Western Kenya tested lime sales incentives for OAF field officers. These incentives were found to increase lime purchases by 13 percentage points. This program involved a payment to field officers of $0.5 per adopting farmer, plus a day of training for the field officers (OAF, 2019). We estimate that the cost per adopting farmer was approximately $2.[42] Therefore, text messages compare relatively

[39] Ideally, one would experimentally estimate the rate of return of these programs to judge whether these programs are worth the investment. However, our experiments were not designed or powered to detect yields impacts. Producing reliable estimates on the returns to text message interventions is difficult since the effects are modest and outcome variables like crop yields and profits tend to be extremely noisy.

[40] We make similar assumptions about the other in-person programs when we make comparisons.

[41] Each event hosted between 100-300 farmers and the overall cost of administering the event was $2,600. This included transport, compensation and materials required to set up the test plots, invite presenters, and carry out the events. In India, farmer field days organized by an NGO were estimated to cost approximately $5 per farmer (Emerick et al., 2016)

[42] In addition to the per farmer adoption incentives, we calculate that the per field officer per season cost to implement this program is at least $20, which includes some training and proportional compensation for additional time on lime sales and transport. Each field officers can then target about 200 farmers.

favorably to these other in-person interventions, especially with bulk texting.

Cost-Benefit. We now provide a cost-benefit approximation of the program. Table 4 reports the parameters utilized in the analysis. To estimate lime benefits, we use the median of four agronomic trials in the region measuring the impact of lime application on maize yields and calculate a 14 kg maize yield increase per 10 kg of lime applied.[43] We estimate that the profits obtained from an additional 10 kg of lime are approximately $3.10, which takes into account the revenue from additional maize sales using prevailing market prices, minus the costs of applying lime (both input costs and labor) and the additional labor costs from harvesting and transport.[44] At the estimated lime application rate (estimated through a meta-analysis across projects), we calculate a benefit-cost ratio of 9:1. However, with at-scale unit costs of $0.001 per text message, the implied benefit cost ratio is closer to 90:1.

For fertilizer, we obtain the impact of 10 kg of application on yields, 24.8 kg, from (Duflo et al., 2011). The cost of applying 10 additional kg of fertilizer is estimated to be approximately $7.4, which takes into account the local price of fertilizer and transport and application cost.[45] The overall impact of the programs in terms of quantity of fertilizer applied implies a profit of $0.03 per farmer treated. Considering that the cost of the programs was on average $0.04 per farmer, the benefit-cost ratio is 0.85. However, at scale, with a unit cost of $0.001 per SMS, the implied benefit-cost ratio would be 9:1. Combining the two components, lime and fertilizer, in a 7-message intervention, we obtain a benefit-cost ratio of 55:1 at scale. These calculations should be interpreted with caution since they rely on several assumptions. However, they are encouraging. First, the estimates on impacts that we use are likely lower bounds. As previously discussed, there is evidence of information sharing among farmers and effect persistence over multiple seasons, which we do not include in these calculations. Second, unlike other in-person programs where treatment costs are likely to rise with wider implementation, operating these programs at scale would significantly reduce costs. Third, organizations can easily choose the most successful approaches for later programs.

[43]We focus on studies that estimated the effect of micro-dosing lime on maize. The impact per 10 kg of lime applied was: 18 kg (OAF, 2014), 25 kg (OAF, 2015), 2 kg (Kisinyo et al., 2015), and 10 kg (Omenyo et al., 2018).

[44]Maize prices and assumed costs are based on data collected by IPA in the study area during between June 2016 and April 2017.The price of lime reflects the average price of lime in western Kenya during the 2018 main agricultural season.

[45]The price of fertilizer reflects the average price of CAN and urea in western Kenya between June 2016 and April 2017, based on data collected by IPA-K.

7 Conclusion

An extensive literature in economics has identified informational barriers as a constraint to behavior change and technology adoption. The rapid uptake of cellphones in developing countries has opened new opportunities to reach people with timely and customized messages. Using text messages to convey information might be a promising tool to reach people at scale, especially in low-income countries, where more intensive or sophisticated approaches remain limited. Yet understanding whether the impacts of these approaches scale to different populations and contexts is critical for policy design. Moreover, if effect sizes are too dependent on the exact implementation features, it might be challenging to know when these interventions will work.

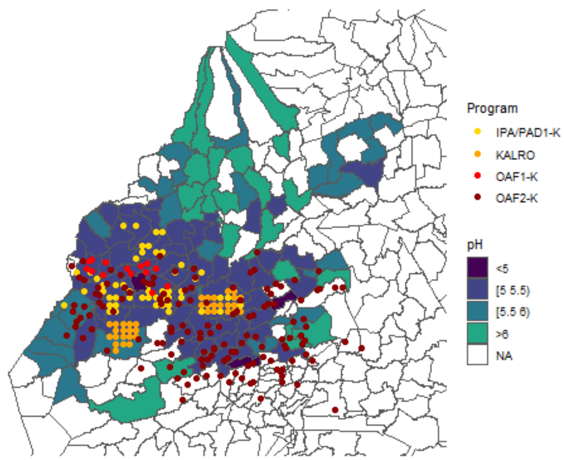
We rigorously evaluated the effects of six different programs implemented in Kenya and Rwanda using actual input purchases as our preferred outcome measure and employing large sample sizes to detect small impacts. Overall, we conclude that the programs had modest but positive effects on the use of agricultural technologies. While it is difficult to make conclusive statements about impact heterogeneity with only six projects, we failed to find strong evidence to support the idea that differences in programs significantly affected impacts. Moreover, our back-of-the-envelope calculations suggest that text-based approaches can be cost-effective from the point of view of an organization that is interested in promoting new inputs. The results highlight the importance of well-powered experiments, especially for very cheap interventions, and caution against making conclusions about the external validity of programs by simply taking non-significant results as evidence of no impact.

While we cannot fully disentangle the mechanisms through which these programs operate, we show that impacts were likely to decay over time, but re-treating farmers sustained the effects. This result suggests that the messages do more than simply create long-lasting knowledge about inputs. If knowledge or awareness were the main channels, one would also expect that the programs would be most effective for those farmers that knew nothing about the new technologies at baseline. Yet, we do not have evidence of differential impacts by baseline knowledge. Moreover, neither providing more information nor adding an in-person phone call significantly changed behavior. On the other hand, against some behavioral predictions, the exact way messages were framed did not significantly increase impacts. Taken together, we conclude that these types of programs can have modest but consistent effects regardless of the exact way in which the programs are designed.

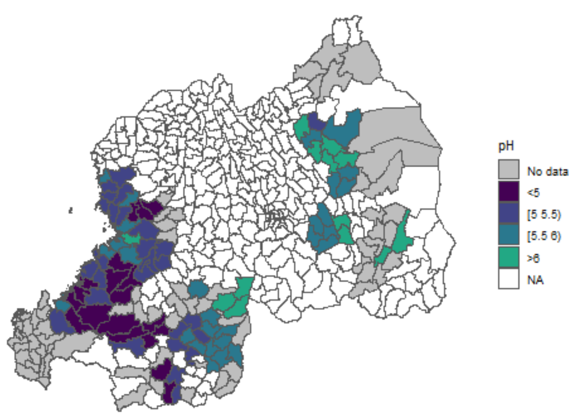
As more sophisticated mobile technologies improve and are adopted over time, more opportunities to better convey information are likely to open up. There is a large scope for policymakers and researchers to continue exploring how to effectively deliver information at scale in cheaper ways.

8 Figures and Tables

Figure 1: Project Maps



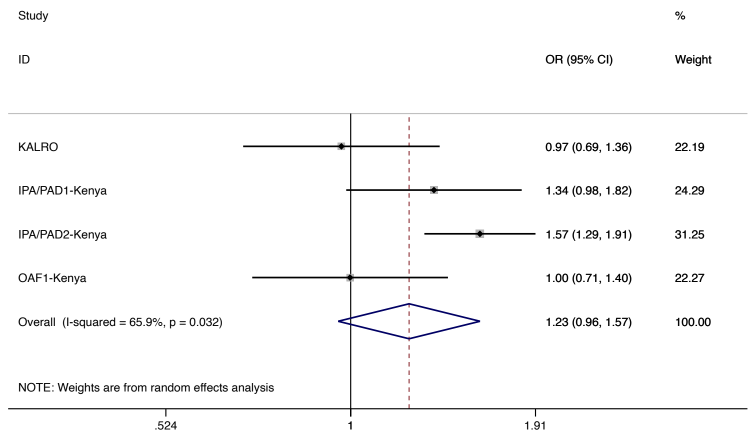
(a) Western Kenya



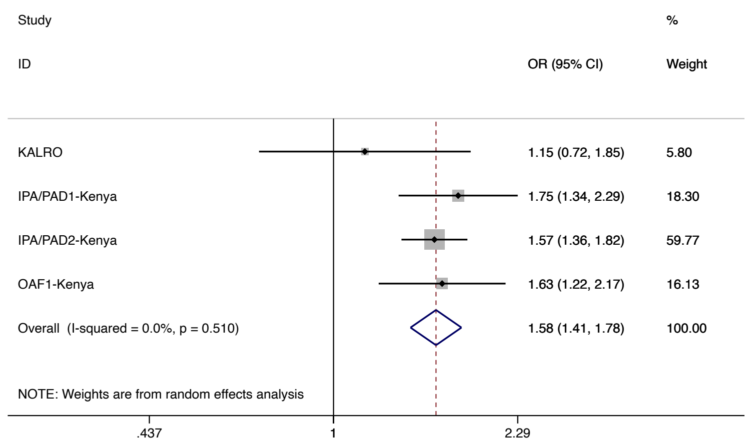
(b) Rwanda

Notes: Panel (a) shows the median level of pH in all wards in which the IPA/PAD2-K program took place as well and the location of the other programs. Panel (b) shows the sectors in which the OAF3-R program took place and the median level of pH, where available.

Figure 2: Effects on Knowledge and Awareness About Lime



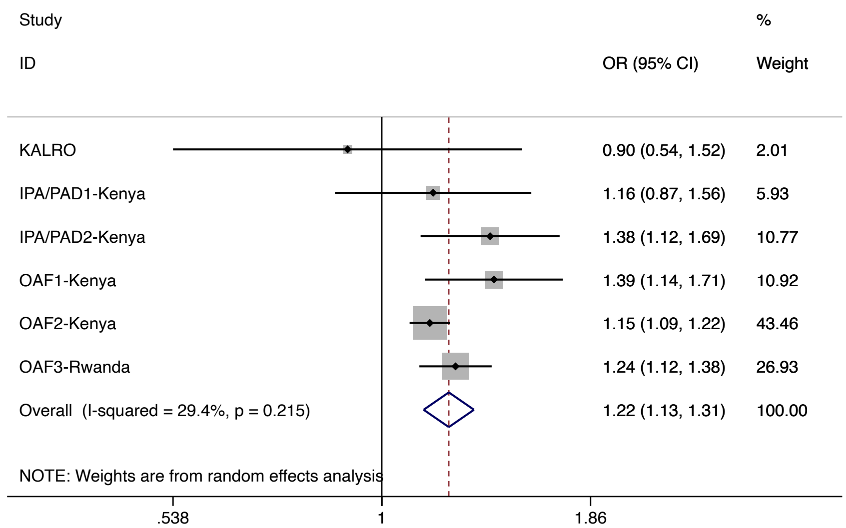
(a) "Have you heard about lime?"



(b) Mentions lime as a way to reduce acidity

Notes: The figure plots the meta-analysis results for specific outcomes. The effects are estimated using a random-effects meta-analysis model. Results are reported in odds ratios. The horizontal lines denote 95% confidence intervals.

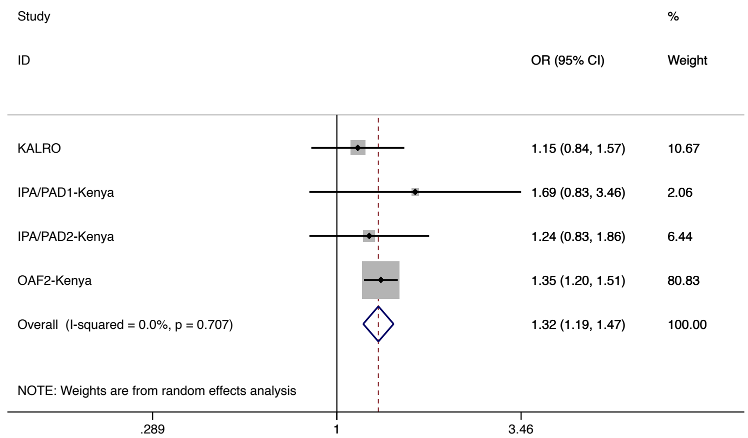
Figure 3: Effects on Lime Purchases (Administrative Data)



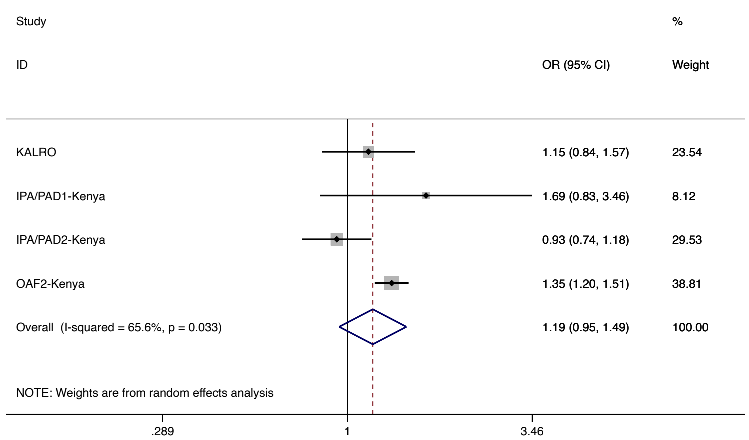
(a) Followed Lime Recommendations

Notes: The figure plots the meta-analysis results for following lime recommendations using administrative data. The effects are estimated using a random-effects meta-analysis model. Results are reported in odds ratios. The horizontal lines denote 95% confidence intervals. The results are measured using administrative data. The KALRO results are measured using coupon redemption in the second season.

Figure 4: Effects on Fertilizer Purchases (Administrative Data)



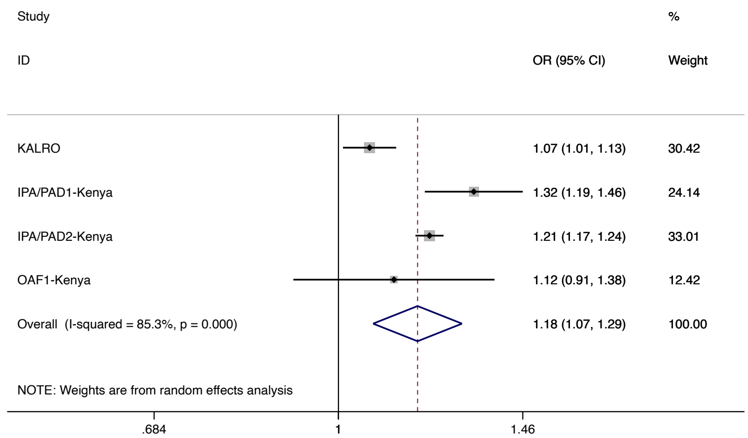
(a) Followed Fertilizer Recommendations



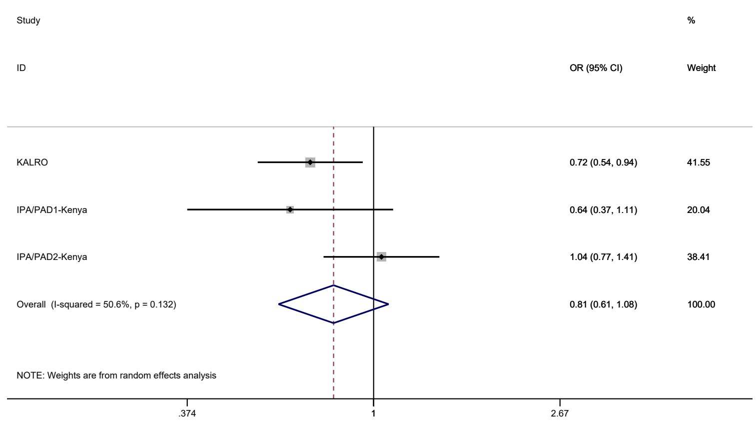
(b) Purchased Any Mentioned Fertilizer

Notes: The figure plots the meta-analysis results for following fertilizer recommendations using administrative data. The effects are estimated using a random-effects meta-analysis model. Results are reported in odds ratios. The horizontal lines denote 95% confidence intervals. All results are measured using administrative data. The KALRO results are measured using coupon redemption in the second season. For Figure (b), the dependent variable for IPA/PAD2-Kenya is a dummy equal to one if either urea or CAN were purchased.

Figure 5: Difference in Survey vs. Administrative Data



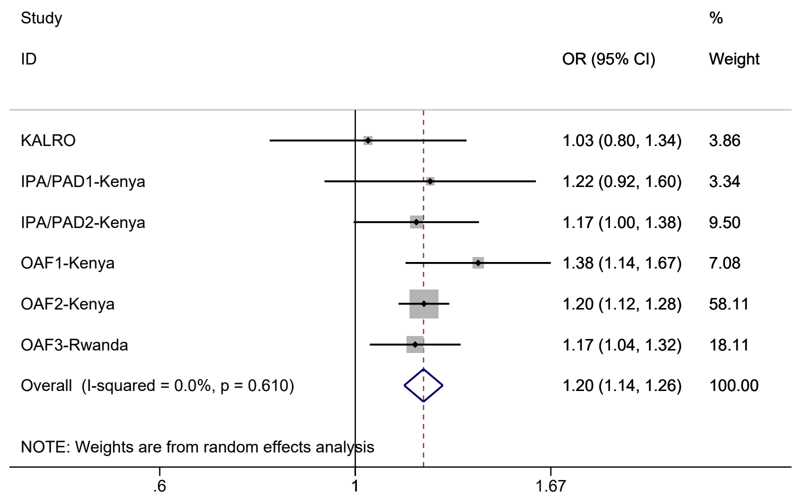
(a) Lime: Ratio Survey/Admin OR



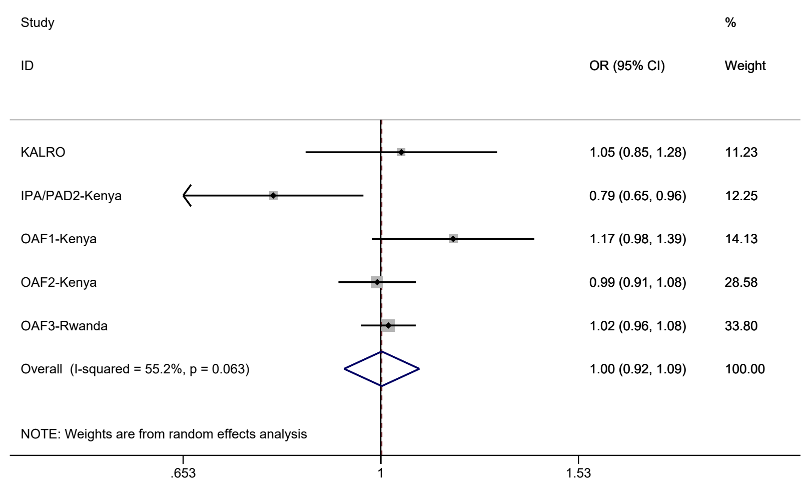
(b) Fertilizer: Ratio Survey/Admin OR

Notes: The figure plots the meta-analysis results for the ratio between the effect of the program on following lime recommendations (figure (a)) and following fertilizer recommendations (figure (b)), measured in terms of odds ratios, estimated using self-report survey data and the same effect estimated using administrative data. The corresponding standard errors are calculated assuming that the correlation between the two estimates is 0.69, which is the correlation between self-reported and administrative data for following lime recommendations for all studies that report both outcomes. The set of studies is restricted to those for which both self-reported and administrative data are available. The combined effects are estimated using a random-effects meta-analysis model. The horizontal lines denote 95% confidence intervals.

Figure 6: Effects on Recommended and Non-Recommended Inputs



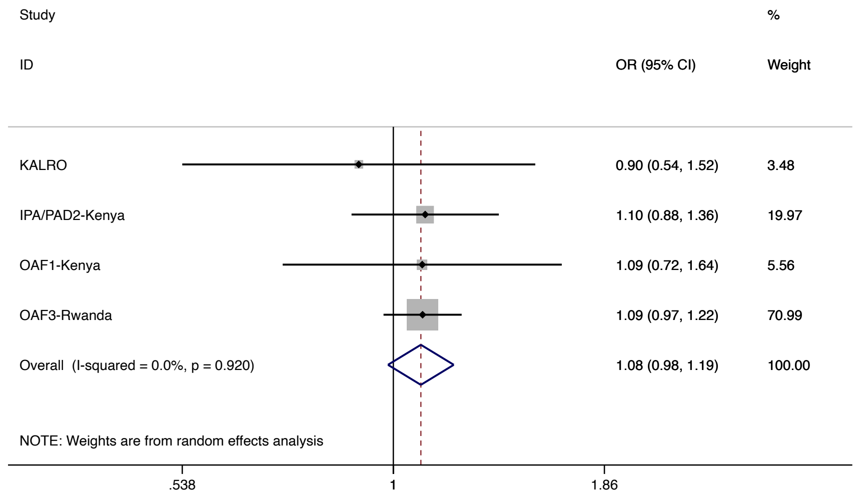
(a) Recommended Inputs



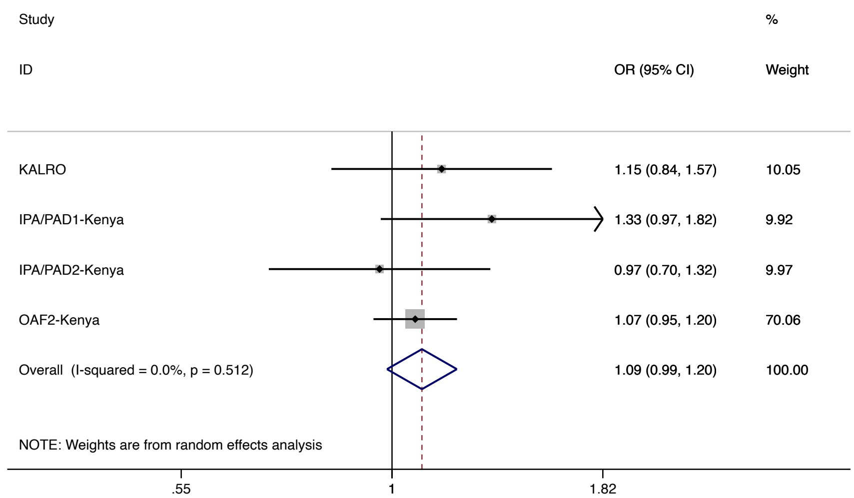
(b) Other Inputs

Notes: The figure plots the meta-analysis results for the effect of the programs on use or purchases of recommended inputs and other inputs not mentioned by the SMS messages. The effects are estimated using a random-effects meta-analysis model. Multiple outcomes per study are aggregated assuming that correlation across outcomes is equal 0.17, which is the correlation between the dummy variables indicating following lime and fertilizer recommendations. Coefficients are obtained from regressions that do not include controls and fixed effects, to ensure convergence. Results are reported in odds ratios. The horizontal lines denote 95% confidence intervals. Panel (a) reports results for recommended inputs. Panel (b) reports results for other inputs. IPA/PAD1-K is not included in (b) because no data for other inputs was collected in that case.

Figure 7: Effect Persistence Over Subsequent Season



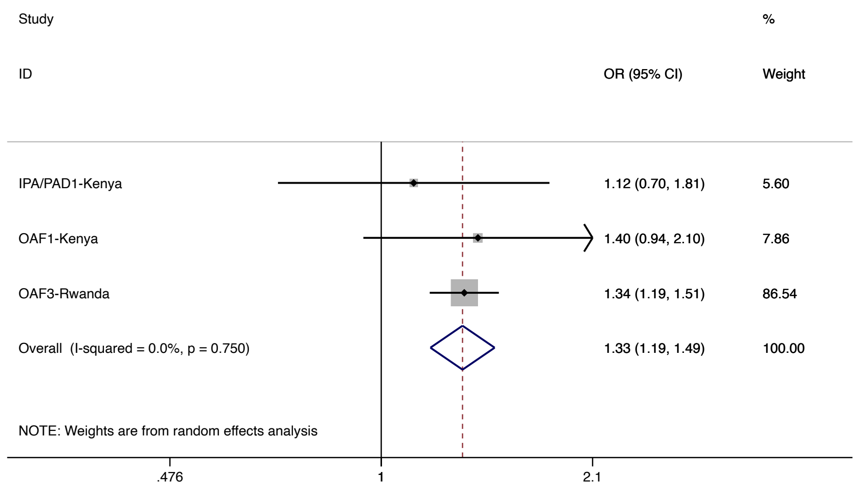
(a) Lime: followed recommendations in second season



(b) Fertilizer: followed recommendations in second season

Notes: The figure plots the meta-analysis results following lime and fertilizer recommendations in the second season (when treated in the first season only). The effects are estimated using a random-effects meta-analysis model. Results are reported in odds ratios. The horizontal lines denote 95% confidence intervals. Panel (a) reports results for following lime recommendations in the second season. Panel (b) reports results for following fertilizer recommendations in the second season.

Figure 8: Message Fatigue



(a) Lime: followed recommendations in second season

Notes: The figure plots the meta-analysis results following lime recommendations in the second season when treated both in the first and subsequent season, compared to the control group. The effects are estimated using a random-effects meta-analysis model. Results are reported in odds ratios. The horizontal lines denote 95% confidence intervals.

Table 3: Summary of Meta-analytic Results

#	Outcome	Effects					Heterogeneity				Pred. Interval 95% CI	
		N	Effect	95% CI		Q stat (p)	I^2	I^2 - 95% CI		τ^2		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Odds Ratios											
1	Heard Lime	4	1.23	0.96	1.57	0.03	65.87	0.00	88.37	0.04	0.44	3.44
2	Knowledge Acidity	4	1.58	1.41	1.78	0.51	0.00	0.00	84.69	0.00	1.23	2.04
3	Lime Rec.	6	1.22	1.13	1.31	0.21	29.37	0.00	71.10	0.00	1.03	1.45
4	Fertilizer Rec.	4	1.32	1.19	1.47	0.71	0.00	0.00	84.69	0.00	1.06	1.66
5	All Recommended Inputs	6	1.20	1.14	1.26	0.61	0.00	0.00	74.62	0.00	1.11	1.28
6	Other Inputs	5	1.00	0.92	1.09	0.06	55.21	0.00	83.46	0.00	0.78	1.28
7	Persistence Lime	4	1.08	0.98	1.19	0.92	0.00	0.00	84.69	0.00	0.88	1.34
8	Fatigue Lime	3	1.33	1.19	1.49	0.75	0.00	0.00	89.60	0.00	0.64	2.77
9	Persistence Fert.	4	1.09	0.99	1.20	0.51	0.00	0.00	84.69	0.00	0.88	1.35

Notes: Meta-analysis results for each outcome reported in the rows. Column (2) -(5) reports results from a random-effects model; Column (6)-(9) reports heterogeneity results. The coefficient represents the estimated summarized effects across studies. Rows 1-9 report results measured in odds ratios. Row 10-11 reports the results for the quantity of lime and fertilizer, measured in Kgs. The results obtained in rows 5 and 6 are obtained assuming correlation across inputs is 0.17, which is the correlation between following lime and fertilizer recommendations across all studies that report these outcomes.

Table 4: Cost-Benefit Analysis Parameters

Benefits	Lime	Fertilizer	Overall
Impact on following recommendation (Table G1)	0.02	0.02	
Impact on application (kg)	1.18	0.33	
Impact of 1 kg application on yield (kg)	1.40	2.48	
Revenue from additional maize kg (US$)			0.34
Cost per 1 kg of input (US$)	0.18	0.74	
Profit per treated farmer (US$)	0.349	0.034	0.383
Program Costs			
Number of text messages	3	4	7
Cost per text message - program (US$)			0.01
Cost per text message - at scale (US$)			0.001
Benefit-Cost Ratio			
Program costs	11.63	0.85	5.47
At scale	116.33	8.51	54.71

Notes: The impact of 1 kg of application on input on yield is estimated based on information available in the literature. Market prices of maize and fertilizer are obtained from a survey of local agricultural supply dealers conducted between June 2016 and April 2017 in western Kenya. Market prices for lime are obtained from a survey of local agricultural supply dealers conducted in western Kenya during the 2018 main agricultural season. To obtain revenues from sales of maize and cost of purchasing inputs we assume a transport cost of $0.05 per kg. The number of text messages is the average number of topic specific messages received by treated farmers.

References

Aker, J. C. (2017). Using digital technology for public service provision in developing countries. *Digital Revolutions in Public Finance*, 201–24.

Aker, J. C., P. Collier, and P. C. Vicente (2017). Is information power? using mobile phones and free newspapers during an election in Mozambique. *Review of Economics and Statistics 99*(2), 185–200.

Aker, J. C. and M. Fafchamps (2015). Mobile phone coverage and producer markets: Evidence from West Africa. *World Bank Economic Review 29*(2), 262–292.

Aker, J. C., I. Ghosh, and J. Burrell (2016). The promise (and pitfalls) of ICT for agriculture initiatives. *Agricultural Economics 47*(S1), 35–48.

Aker, J. C., C. Ksoll, and T. J. Lybbert (2012). Can mobile phones improve learning? evidence from a field experiment in Niger. *American Economic Journal: Applied Economics 4*(4), 94–120.

Aker, J. C. and I. M. Mbiti (2010, September). Mobile phones and economic development in Africa. *Journal of Economic Perspectives 24*(3), 207–32.

Allcott, H. (2015). Site selection bias in program evaluation. *The Quarterly Journal of Economics 130*(3), 1117–1165.

Angrist, N., P. Bergman, and M. Matsheng (2020). School's out: Experimental evidence on limiting learning loss using "low-tech" in a pandemic. Technical report, National Bureau of Economic Research.

Arouna, A., J. D. Michler, W. G. Yergo, and K. Saito (2019). One size does not fit all: Experimental evidence on the digital delivery of personalized extension advice in nigeria.

Baumüller, H. (2018). The little we know: an exploratory literature review on the utility of mobile phone-enabled services for smallholder farmers. *Journal of International Development 30*(1), 134–154.

Benartzi, S., J. Beshears, K. L. Milkman, C. R. Sunstein, R. H. Thaler, M. Shankar, W. Tucker-Ray, W. J. Congdon, and S. Galing (2017). Should governments invest more in nudging? *Psychological science 28*(8), 1041–1055.

BenYishay, A. and A. M. Mobarak (2013). Communicating with farmers through social networks.

Bird, K. A., B. L. Castleman, J. T. Denning, J. Goodman, C. Lamberton, and K. O. Rosinger (2019). Nudging at scale: Experimental evidence from fafsa completion campaigns. *Journal of Economic Behavior & Organization 183*, 105–128.

Borenstein, M., L. Hedges, J. Higgins, and R. HR (2009). *Introduction to Meta-Analysis*. John Wiley.

Borenstein, M., J. P. Higgins, L. V. Hedges, and H. R. Rothstein (2017a). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research synthesis methods 8*(1), 5–18.

Borenstein, M., J. P. Higgins, L. V. Hedges, and H. R. Rothstein (2017b). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research synthesis methods 8*(1), 5–18.

Brady, N. and R. Weil (2004). Soil acidity, alkalinity and salinity. *Elements of the Nature and Properties of Soils*, 266–315.

Buntaine, M. T., R. Jablonski, D. L. Nielson, and P. M. Pickering (2018). Sms texts on corruption help ugandan voters hold elected councillors accountable at the polls. *Proceedings of the National Academy of Sciences 115*(26), 6668–6673.

Bursztyn, L., S. Fiorin, D. Gottlieb, and M. Kanz (2019). Moral incentives in credit card debt repayment: evidence from a field experiment. *Journal of Political Economy 127*(4), 1641–1683.

Camacho, A. and E. Conover (2010). The impact of receiving price and climate information in the agricultural sector. *Working Paper*.

Casaburi, L., M. Kremer, S. Mullainathan, and R. Ramrattan (2014). Harnessing ICT to increase agricultural production: Evidence from Kenya. *Working Paper*.

Chong, A., D. Karlan, J. Shapiro, and J. Zinman (2015). (ineffective) messages to encourage recycling: evidence from a randomized evaluation in peru. *The World Bank Economic Review 29*(1), 180–206.

Chuang, E., P. Dupas, E. Huillery, and J. Seban (2020). Sex, lies, and measurement: Consistency tests for indirect response survey methods. *Journal of Development Economics 148*, 102582.

Cole, S. A. and A. N. Fernando (2021). 'mobile'izing agricultural advice technology adoption diffusion and sustainability. *The Economic Journal 131*(633), 192–219.

Courtois, P. and J. Subervie (2014). Farmer bargaining power and market information services. *American Journal of Agricultural Economics 97*(3), 953–977.

Cunha, N., G. Lichand, R. Madeira, and E. Bettinger (2017). What is it about communicating with parents? *Working Paper*.

Dahal, M. and N. Fiala (2020). What do we know about the impact of microfinance? the problems of statistical power and precision. *World Development 128*, 104773.

Deichmann, U., A. Goyal, and D. Mishra (2016). Will digital technologies transform agriculture in developing countries? *Agricultural Economics 47*(S1), 21–33.

DellaVigna, S. and E. Linos (2020). Rcts to scale: Comprehensive evidence from two nudge units. Technical report, Working Paper, UC Berkeley.

DerSimonian, R. and N. Laird (1986). Meta-analysis in clinical trials. *Controlled clinical trials 7*(3), 177–188.

Duflo, E., M. Kremer, and J. Robinson (2008). How high are rates of return to fertilizer? evidence from field experiments in Kenya. *The American Economic Review 98*(2), 482–488.

Duflo, E., M. Kremer, and J. Robinson (2011). Nudging farmers to use fertilizer: Theory and experimental evidence from kenya. *The American Economic Review 101*(6), 2350–2390.

Duflo, E., M. Kremer, J. Robinson, and F. Schilbach (2018). Learning about appropriate technology use: Evidence from fertilizer use in western Kenya. *Working Paper*.

Dustan, A., S. Maldonado, and J. M. Hernandez-Agramonte (2018). Motivating bureaucrats with non-monetary incentives when state capacity is weak: Evidence from large-scale field experiments in peru. *Working Paper*.

Emerick, K., A. de Janvry, E. Sadoulet, and M. H. Dar (2016). Identifying early adopters, enhancing learning, and the diffusion of agricultural technology.

Fabregas, R., M. Kremer, J. Robinson, and F. Schilbach (2017a). The effectiveness of public agricultural extension: Evidence from two approaches in Kenya. *Working Paper*.

Fabregas, R., M. Kremer, J. Robinson, and F. Schilbach (2017b). Evaluating agricultural information dissemination in western kenya (3ie impact evaluation report no. 67). *New Delhi: International Initiative for Impact Evaluation*.

Fabregas, R., M. Kremer, and F. Schilbach (2019). Realizing the potential of digital development: The case of agricultural advice. *Science 366*(6471).

Fafchamps, M. and B. Minten (2012). Impact of sms-based agricultural information on Indian farmers. *The World Bank Economic Review 26*(3), 383–414.

FAO (2015). Status of the world's soil resources (swsr)–main report. *Food and agriculture organization of the United Nations and intergovernmental technical panel on soils, Rome, Italy 650*.

Friedman, W., B. Woodman, and M. Chatterji (2015). Can mobile phone messages to drug sellers improve treatment of childhood diarrhoea?—a randomized controlled trial in ghana. *Health policy and planning 30*(suppl_1), i82–i92.

FURP (1995). New fertilizer recommendations for small-scale farmers in kenya. *KARI Information Bulletin - No. 12*.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian data analysis*. CRC press.

Glewwe, P., M. Kremer, S. Moulin, and E. Zitzewitz (2004). Retrospective vs. prospective analyses of school inputs: the case of flip charts in kenya. *Journal of development Economics 74*(1), 251–268.

Grossman, G., M. Humphreys, and G. Sacramone-Lutz (2020). Information technology and political engagement: Mixed evidence from uganda. *The Journal of Politics 82*(4), 1321–1336.

GSMA (2020). 10 years of mobile for development. Technical report.

Gudu, S., J. Okalebo, C. Othieno, P. Obura, D. Ligeyo, D. Shulze, and C. Johston (2005). Response of five maize genotypes to nitrogen, phosphorus and lime on acid soils of Western Kenya. In *African Crop Science Conference Proceedings*, Volume 7, pp. 1109–1115.

Gupta, A., J. Ponticelli, and A. Tesei (2020). Information, technology adoption and productivity: The role of mobile phones in agriculture. Technical report, National Bureau of Economic Research.

Haaland, I., C. Roth, and J. Wohlfart (2020). Designing information provision experiments.

Hall, A. K., H. Cole-Lewis, and J. M. Bernhardt (2015). Mobile text messaging for health: a systematic review of reviews. *Annual Review of Public Health 36*, 393–415.

Hall, C. S., E. Fottrell, S. Wilkinson, and P. Byass (2014). Assessing the impact of mhealth interventions in low-and middle-income countries–what has been shown to work? *Global health action 7*(1), 25606.

Higgins, J. P., S. Green, et al. (2008). Cochrane handbook for systematic reviews of interventions.

Higgins, J. P., J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, and V. A. Welch (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Higgins, J. P. and S. G. Thompson (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine 21*(11), 1539–1558.

Higgins, J. P., S. G. Thompson, J. J. Deeks, and D. G. Altman (2003). Measuring inconsistency in meta-analyses. *Bmj 327*(7414), 557–560.

Ioannidis, J., T. Stanley, and H. Doucouliagos (2017). The power of bias in economics research. *The Economic Journal 127*(605), F236–F265.

IPNI (1999). Nitrogen notes. *Manual, http://www.ipni.net/NitrogenNotes*.

Jamison, J. C., D. Karlan, and P. Raffler (2013). Mixed-method evaluation of a passive mhealth sexual information texting service in uganda. *Information Technologies & International Development 9*(3), pp–1.

Jensen, R. (2007). The digital provide: Information (technology), market performance, and welfare in the south Indian fisheries sector. *The Quarterly Journal of Economics*.

Karlan, D., M. McConnell, S. Mullainathan, and J. Zinman (2016). Getting to the top of mind: How reminders increase saving. *Management Science 62*(12), 3393–3411.

Karlan, D., M. Morten, and J. Zinman (2012). A personal touch: Text messaging for loan repayment. Technical report, National Bureau of Economic Research.

Karlan, D., M. Morten, and J. Zinman (2015). A personal touch in text messaging can improve microloan repayment. *Behavioral Science & Policy 1*(2), 25–31.

Karlan, D. and J. Zinman (2008). Lying about borrowing. *Journal of the European Economic Association 6*(2-3), 510–521.

Karlan, D. S. and J. Zinman (2012). List randomization for sensitive behavior: An application for measuring use of loan proceeds. *Journal of Development Economics 98*(1), 71–75.

Kelly, V. and A. Murekezi (2000). Fertilizer response and profitability in rwanda. *A Synthesis of Findings from MINAGRI Studies Conducted by the Food Security Research Project (FSRP) and the FAO Soil Fertility Initiative. Kigali, Rwanda: FSRP/MINAGRI, February*.

Kisinyo, P., P. Opala, V. Palapala, S. Gudu, C. Othieno, and E. Ouma (2015). Micro-dosing of lime, phosphorus and nitrogen fertilizers effect on maize performance on an acid soil in Kenya. *Sustainable Agriculture Research 4*(2), 21.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

KSHC (2014). Policy brief: Soil health research gaps in kenya. https://www.kalro.org/sites/default/files/KSHC

Larochelle, C., J. Alwang, E. Travis, V. H. Barrera, and J. M. D. Andrade (2019). Did you really get the message? Using text reminders to stimulate adoption of agricultural technologies. *The Journal of Development Studies 55*(4), 548–564.

Linos, E., A. C. Prohofsky, A. Ramesh, J. Rothstein, and M. Unrath (2020). Can nudges increase take-up of the eitc?: Evidence from multiple field experiments. *NBER Working Paper* (w28086).

Lobell, D. B., G. Azzari, M. Burke, S. Gourlay, Z. Jin, T. Kilic, and S. Murray (2018). *Eyes in the sky, boots on the ground: assessing satellite-and ground-based approaches to crop yield measurement and analysis in Uganda*. The World Bank.

McKenzie, D. and C. Woodruff (2014). What are we learning from business training and entrepreneurship evaluations around the developing world? *The World Bank Research Observer 29*(1), 48–82.

Meager, R. (2019). Understanding the average impact of microcredit expansions: A bayesian hierarchical analysis of seven randomized experiments. *American Economic Journal: Applied Economics 11*(1), 57–91.

Mitra, S., D. Mookherjee, M. Torero, and S. Visaria (2017). Asymmetric information and middleman margins: An experiment with Indian potato farmers. *Review of Economics and Statistics* (0).

Mortvedt, J. L. M. and R. Follet (1999). Fertilizer technology and application. *Meister Publishing Co. Willoughby, OH*.

NAAIAP (2014). Soil suitability evaluation for maize production in kenya. https://africafertilizer.org/wp-content/uploads/2017/04/Soil-Suitability-Evaluation-for-Maize-Prod.-in-Kenya.pdf.

Nakasone, E., M. Torero, and B. Minten (2014). The power of information: The ICT revolution in agricultural development. *Annu. Rev. Resour. Econ. 6*(1), 533–550.

OAF (2014). Managing soil acidity with lime (2014 phase 2 trial report). Technical report, One Acre Fund.

OAF (2015). Managing soil acidity with lime (2015 trial report). Technical report, One Acre Fund.

OAF (2017). Annual report.

OAF (2019). Lime experience trial results. *One Acre Fund Report*.

Omenyo, V. S., J. Okalebo, and C. Othieno (2018). Effects of lime and phosphorus fertilizer on maize performance in acid soils of western Kenya. *Second RUFORUM Biennial Meeting 20 - 24 September 2010, Entebbe, Uganda*.

Oreopoulos, P. and U. Petronijevic (2019). The remarkable unresponsiveness of college students to nudging and what we can learn from it. Technical report, National Bureau of Economic Research.

Plaster, J. (2003). Soil science and management. *Delmar Learning, a division of Thomas Learning, Inc*, 319–322.

Pop-Eleches, C., H. Thirumurthy, J. P. Habyarimana, J. G. Zivin, M. P. Goldstein, D. De Walque, L. Mackeen, J. Haberer, S. Kimaiyo, J. Sidle, et al. (2011). Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. *AIDS (London, England) 25*(6), 825.

Pritchett, L. and J. Sandefur (2015). Learning from experiments when context matters. *American Economic Review 105*(5), 471–75.

Raifman, J. R., H. E. Lanthorn, S. Rokicki, and G. Fink (2014). The impact of text message reminders on adherence to antimalarial treatment in northern ghana: a randomized trial. *PloS one 9*(10), e109032.

Rokicki, S., J. Cohen, J. A. Salomon, and G. Fink (2017). Impact of a text-messaging program on adolescent reproductive health: a cluster–randomized trial in ghana. *American journal of public health 107*(2), 298–305.

Rosenzweig, M. R. and C. Udry (2020). External validity in a stochastic world: Evidence from low-income countries. *The Review of Economic Studies 87*(1), 343–381.

Rubin, D. (1981). Estimation in parallel randomized experiments. *Journal of Educational and Behavioral Statistics*.

Steinhardt, L. C., D. P. Mathanga, D. Mwandama, H. Nsona, D. Moyo, A. Gumbo, M. Kobayashi, R. Namuyinga, M. P. Shah, A. Bauleni, et al. (2019). The effect of text message reminders to health workers on quality of care for malaria, pneumonia, and diarrhea in malawi: a cluster-randomized, controlled trial. *The American journal of tropical medicine and hygiene 100*(2), 460–469.

Svensson, J. and D. Yanagizawa (2009). Getting prices right: the impact of the market information service in uganda. *Journal of the European Economic Association 7*(2-3), 435–445.

Terman, G. and O. Engelstad (1976). Agronomic evaluation of fertilizers: Principles and practices. *Muscle Shoals, Ala.: Tennessee Valley Authority*.

The, C., H. Calba, E. L. M. Zonkeng, C.and Ngonkeu, and V. O. Adetimirin (2006). Response of maize grain yield to changes in acid soil characteristics after soil amendment. *Plant Soil*.

Tisdale, S. L., W. L. Nelson, and J. D. Beaton (1990). Soil fertility and fertilizers. *New York: Macmillan*.

Tjernström, E., T. J. Lybbert, R. F. Hernández, and J. S. Correa (2019). Learning by (virtually) doing: experimentation and belief updating in smallholder agriculture. Technical report, Working Paper, University of Wisconsin-Madison.

Van Campenhout, B., D. J. Spielman, and E. Lecoutere (2020). Information and communication technologies to provide agricultural advice to smallholder farmers: Experimental evidence from uganda. *American Journal of Agricultural Economics*.

Vivalt, E. (2016). How much can we generalize from impact evaluations? *Working Paper*.

Wasonga, C. J., D. O. Sigunga, and A. O. Musundu (2008). Phosphorus requirements by maize varieties in different soil types of Western Kenya. *Africa Crop Science Journal vol. 16*, 161–173.

Wiecek, W. and R. Meager (2022). baggr: Bayesian aggregation package for r, v0.6.16. https://github.com/wwiecek/baggr.

Appendix

Table of Contents

A Regression Controls and Variables

Table A1: Inputs Measured and Control Variables

Sample	Recommended Inputs	Non-Recommended Inputs Other	Non-Recommended Inputs Fertilizers	Control Variables (baseline)
KALRO	Lime, planting fertilizer (DAP, NPK), top-dressing fertilizer (CAN, Mavuno), compost, manure, hybrid seeds	Rhizobia, striga control, pest and disease control, storage bags		female, lime awareness, input use index, grow legumes, land size, soil test knowledge
IPA/PAD1-K	Lime, DAP, urea		NPK, CAN, Mavuno	age,female, education, database type, language, darm size, phone network, knowledge, input use, interest in program
IPA/PAD2-K	Lime, DAP, urea	hybrid seeds, pesticides	NPK, CAN, Mavuno	female, age, language, land size, input use, agrovet recruiter dummies
OAF1-K	Lime	Actellic, compost, extra CAN, drying sheets, storage bags, machete, hoe		seasons in OAF, group size, repayment incentive, prior orders of: maize package, bean seeds, compost products, solar lamps, cookstoves extra CAN, harvest sheets, storage bags, onion seeds health insurance sanitary pads
OAF2-K	Lime, extra CAN	Actellic, compost, drying sheets, storage bags		seasons in OAF, group size, predicted pH, prior orders of: maize package, solar lamps, extra CAN health insurance sanitary pads
OAF3-R	Lime	DAP, NPK, urea, storage bags		seasons in OAF, group size, predicted pH, prior orders of: fertilizer (DAP and NPK), lime, urea credit size

Notes: The table shows the list of recommended inputs used for which we have administrative or survey data at endline, the list of non-recommended inputs for which we have survey data at endline, and the list of control variables included in our main regressions for each experiment (all of them measuring prior to the program introduction). Included controls are constrained by data availability for each project. Results are very similar when controls are excluded.

B Attrition & Balance

Table B1: KALRO: Summary Statistics & Balance

	Control	Treated	(1) vs. (2)
	(1)	(2)	(3)
Age	41.33	39.79	1.54*
	(0.66)	(0.65)	(0.92)
Female	0.64	0.65	-0.01
	(0.02)	(0.02)	(0.03)
Primary school	0.53	0.54	-0.02
	(0.02)	(0.02)	(0.03)
Secondary school	0.03	0.04	-0.01
	(0.01)	(0.01)	(0.01)
Footwear	0.61	0.56	0.05
	(0.02)	(0.02)	(0.03)
Mumias	0.56	0.57	-0.01
	(0.02)	(0.02)	(0.03)
Acres (owned and rented)	2.22	1.92	0.29
	(0.26)	(0.10)	(0.28)
Literate	0.91	0.91	0.00
	(0.01)	(0.01)	(0.02)
Had soil test	0.12	0.10	0.02
	(0.02)	(0.01)	(0.02)
Mentions Lime	0.03	0.05	-0.02
	(0.01)	(0.01)	(0.01)
Used Lime	0.06	0.07	-0.01
	(0.01)	(0.01)	(0.02)
Used fertilizer last LR season	0.84	0.84	0.00
	(0.02)	(0.02)	(0.03)
Grows legumes	0.81	0.83	-0.03
	(0.02)	(0.02)	(0.03)
Heard Lime	0.39	0.40	0.00
	(0.02)	(0.02)	(0.03)
Heard soil test	0.80	0.87	-0.07**
	(0.02)	(0.02)	(0.03)
Ever used DAP	0.94	0.94	0.00
	(0.01)	(0.01)	(0.02)
Ever used CAN	0.61	0.63	-0.03
	(0.02)	(0.02)	(0.03)
Ever used NPK	0.12	0.14	-0.02
	(0.02)	(0.02)	(0.02)
N	418	415	833
Joint F-Stat			1.06
P-value			0.386

Notes: The table shows summary statistics and balance tests using covariate variables from a baseline survey. Columns (1)–(2) display the mean and standard error of each characteristic for each treatment group. Column (3) displays the differences across columns and corresponding standard error. *Mumias* denotes share of farmers from Kakamega county (Mumias area), *Had soil test* denotes ever having a soil test, *Mentions Lime* is a dummy variable with value one if respondent mentioned lime as a strategy to reduce soil acidity. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table B2: IPA/PAD1-K: Summary Statistics & Balance

	Control (1)	General (2)	Specific (3)	(1) vs. (2) (4)	(1) vs. (3) (5)	(2) vs. (3) (6)
Age	46.25	46.01	45.59	0.25	0.66	0.42
	(0.49)	(0.45)	(0.43)	(0.66)	(0.65)	(0.63)
Female	0.37	0.37	0.37	-0.01	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Primary school	0.60	0.61	0.66	-0.01	-0.05*	-0.04
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Secondary school	0.10	0.10	0.10	0.00	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Mumias	0.53	0.53	0.53	0.00	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
pH prediction	5.42	5.40	5.40	0.02	0.02	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Prefers English	0.30	0.27	0.30	0.03	0.00	-0.03
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Heard Lime	0.16	0.17	0.17	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Acres (owned and rented)	2.00	1.86	2.14	0.14	-0.14	-0.28
	(0.09)	(0.08)	(0.31)	(0.12)	(0.32)	(0.32)
Used Lime	0.12	0.13	0.12	-0.01	0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Used DAP last LR season	0.78	0.78	0.80	0.00	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Used NPK last LR season	0.04	0.05	0.04	-0.01	0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Used CAN last LR season	0.62	0.62	0.59	0.00	0.02	0.02
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Used Urea last LR season	0.18	0.18	0.18	0.00	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Used Mavuno last LR season	0.15	0.13	0.16	0.02	-0.01	-0.03
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Main network	0.95	0.94	0.94	0.01	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
N	632	633	632	1265	1264	1265
Joint F-Stat				0.57	0.78	0.75
P-value				0.909	0.714	0.746

Notes: The table shows summary statistics and balance tests using covariate variables from a baseline survey. Columns (1)–(3) display the mean and standard error of each characteristic for each treatment group. Columns (4)-(6) display the difference across columns and the corresponding standard error. *MSC Sample* denotes share of farmers from the Mumias Sugar Company sample. *pH prediction* represents the median pH level measured in the farmer's catchment area. *Mentions Lime* is a dummy variable with value one if the respondent mentioned lime as a strategy to reduce soil acidity. Fertilizer use variables refer to input use during the 2016 long rain season. *Main network* indicates whether the farmer's phone service provider is the main network in the area. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table B3: IPA/PAD2-K: Additional Summary Statistics & Balance

	Control	SMS	SMS+Call	SMS+Call Offer	(1) vs. (2)	(1) vs. (3)	(1) vs. (4)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	42.10	41.40	41.48	41.44	0.70	0.61	0.66
	(0.32)	(0.31)	(0.32)	(0.31)	(0.45)	(0.46)	(0.45)
Female	0.34	0.34	0.34	0.34	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Primary school	0.72	0.70	0.69	0.71	0.01	0.02	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Secondary school	0.13	0.13	0.12	0.13	-0.01	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
pH prediction	5.37	5.37	5.37	5.37	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Prefers English	0.36	0.35	0.34	0.35	0.01	0.02	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Heard Lime	0.26	0.26	0.24	0.25	0.00	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Acres (owned and rented)	2.02	1.85	2.09	2.03	0.17**	-0.07	-0.02
	(0.06)	(0.05)	(0.09)	(0.06)	(0.08)	(0.11)	(0.08)
Maize yield (t/ha)	1.51	1.46	1.37	1.49	0.05	0.15***	0.02
	(0.04)	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)	(0.05)
Used Lime	0.09	0.09	0.09	0.10	-0.01	0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
OAF Participant	0.35	0.34	0.35	0.37	0.00	-0.01	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Used CAN last LR season	0.64	0.62	0.65	0.62	0.02	0.00	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Used Urea last LR season	0.18	0.20	0.20	0.18	-0.02	-0.02	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Used Mavuno last LR season	0.08	0.08	0.07	0.09	0.00	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Lime rec	0.77	0.76	0.77	0.76	0.01	0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
N	1470	1475	1473	1472	2945	2943	2942
Joint F-Stat					0.93	0.64	0.51
P-value					0.52	0.84	0.93

Notes: The table shows summary statistics and balance tests using covariate variables from a baseline survey. Columns (1)–(4) display the mean and standard error of each characteristic for each treatment group. Columns (5)-(7) display the difference across columns and the corresponding standard error. *pH prediction* represents the median pH level measured in the farmer's ward used to provide lime recommendations. *OAF Participant* is dummy variable indicating whether the farmer has ever been enrolled in the OAF program. *Mentions Lime* is a dummy variable with value one if the respondent mentioned lime as a strategy to reduce soil acidity. Fertilizer use variables refer to input use during the 2016 long rains season. *Recommended lime* indicates whether the farmer resided in a ward where lime was recommended. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table B4: OAF1-K: Additional Summary Statistics & Balance

	Broad (1)	Control (2)	Detailed (3)	(1) vs. (2) (4)	(1) vs. (3) (5)	(2) vs. (3) (6)
Female	0.64	0.64	0.67	0.00	-0.02	-0.03
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Group size	9.24	9.08	9.07	0.16	0.17*	0.01
	(0.07)	(0.07)	(0.07)	(0.10)	(0.10)	(0.10)
OAF Seasons	1.50	1.51	1.52	0.00	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Maize inputs (acres)	0.49	0.50	0.50	-0.01	-0.02	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Repayment Incentive (hoe)	0.07	0.06	0.08	0.01	-0.02**	-0.02***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
pH prediction	5.48	5.48	5.48	0.00	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Intercropped (acres)	0.01	0.01	0.01	0.00*	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Extra CAN purchased	0.05	0.04	0.05	0.01**	0.00	-0.02***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Onions (quantity)	0.13	0.09	0.12	0.04***	0.01	-0.03**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Storage Bags	0.31	0.23	0.24	0.07*	0.06	-0.01
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Solar Lamps	0.45	0.44	0.46	0.01	-0.01	-0.02
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Health Insurance	0.23	0.22	0.21	0.01	0.02	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
N	1684	1559	1641	3243	3325	3200
Joint F-Stat				1.74	1.93	1.71
P-value				0.053	0.026	0.058

Notes: The table shows summary statistics and balance tests using covariate variables from OAF long rain 2016 administrative records (before the trial took place). Columns (1)-(3) display mean and standard errors of each variable, by treatment group. Columns (4)-(6) display the difference across columns and the corresponding standard error. *Group size* denotes the number of farmers in the participant's OAF group, *OAF seasons* denotes the number of seasons of enrollment in the OAF program, *Maize inputs (acres)* represents the size of the maize inputs package purchased, *Repayment Incentive* is a dummy variable with value one if the farmer obtained a hoe as bonus for early repayment, *pH prediction* is the variable obtained using kriging interpolation that was used to produce detailed recommendations. *Intercropped* indicates the size of the beans input package, for maize-beans intercropping, *Extra CAN*, *Onions*, *Solar Lamps*, and *Health Insurance* are dummy variables equal to one if the farmer purchased those additional products. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table B5: OAF2-K: Summary Statistics & Balance

	Control	Lime + CAN	Lime only	(1) vs. (2)	(1) vs. (3)	(2) vs. (3)
	(1)	(2)	(3)	(4)	(5)	(6)
Age	48.40	48.44	48.30	-0.04	0.10	0.14
	(0.15)	(0.20)	(0.10)	(0.25)	(0.18)	(0.22)
Female	0.69	0.68	0.69	0.01	0.00	-0.01
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Group size	9.87	9.92	9.82	-0.05	0.04	0.10**
	(0.03)	(0.04)	(0.02)	(0.05)	(0.04)	(0.05)
OAF Seasons	2.23	2.22	2.23	0.01	0.00	-0.01
	(0.02)	(0.02)	(0.01)	(0.03)	(0.02)	(0.02)
Maize inputs (acres)	0.51	0.53	0.51	-0.01**	0.00	0.01**
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)
pH prediction	5.33	5.33	5.33	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Intercropped (acres)	0.01	0.01	0.01	0.00	0.00*	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Extra CAN purchased	0.15	0.14	0.15	0.01	0.00	-0.01
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
Onions	0.04	0.04	0.04	-0.01	0.00	0.01*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Storage Bags	0.40	0.40	0.42	0.00	-0.02	-0.02
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Solar Lamp	0.42	0.43	0.42	0.00	0.00	0.01
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Credit size	9504.07	9616.95	9467.78	-112.88	36.29	149.17**
	(49.84)	(64.63)	(31.67)	(81.55)	(58.68)	(71.17)
N	8142	4872	19558	13014	27700	24430
Joint F-Stat				0.63	0.86	1.78
P-value				0.816	0.583	0.045

Notes: The table shows summary statistics and balance tests using covariate variables from OAF long rain 2017 administrative records (before the trial took place). Columns (1)-(3) display mean and standard errors of each variable, by treatment group. Columns (4)-(6) display the difference across columns and the corresponding standard error. *Group size* denotes the number of farmers in the participant's OAF group, *OAF seasons* denotes the number of seasons of enrollment in the OAF program, *Maize inputs (acres)* represents the size of the maize inputs package purchased, *pH prediction* was obtained using kriging interpolation. *Intercropped* indicates the size of the beans input package, for maize-beans intercropping, *Extra CAN*, *Onions*, *Solar Lamps*, are dummy variables equal to one if the farmer purchased those additional products. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table B6: OAF3-R: Summary Statistics & Balance

	Full Control	Full Treatment	Partial Treatment		(1) vs. (2)	(1) vs. (3)	(1) vs. (4)
			Non Treated	Treated			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Group size	10.73	10.75	10.74	10.73	-0.02	-0.01	0.00
	(0.06)	(0.05)	(0.04)	(0.04)	(0.08)	(0.07)	(0.07)
OAF Seasons	2.01	2.02	2.02	2.01	-0.00	-0.00	-0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Bought lime	0.06	0.06	0.06	0.06	0.00	0.01*	0.01*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Planting fertilizer (kg)	13.78	13.84	13.74	13.80	-0.06	0.04	-0.02
	(0.20)	(0.14)	(0.13)	(0.14)	(0.24)	(0.24)	(0.24)
Bought urea	0.75	0.75	0.74	0.75	-0.00	0.01	-0.00
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Solar Lamp	0.28	0.27	0.28	0.28	0.01	-0.01	-0.00
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Credit size	22768.25	22564.28	22983.56	22941.68	203.97	-215.31	-173.43
	(213.22)	(154.89)	(157.12)	(157.15)	(263.52)	(264.84)	(264.86)
N	19743	37809	28520	28497	57552	48263	48240
Joint F-Stat					0.26	0.81	0.55
P-value					0.97	0.58	0.80

Notes: The table shows summary statistics and balance tests using covariate variables from OAF 2016 administrative records (before the trial took place). Columns (1) - (4) display mean and standard errors of each variable, by treatment group. Columns (5)-(7) displays the difference across columns and the corresponding standard error. *Group size* denotes the number of farmers in the participant's OAF group, *OAF seasons* denotes the number of seasons of enrollment in the OAF program, *Bought lime* is a dummy indicating whether the farmer purchased lime. *Planting fertilizer* indicates the quantity of planting fertilizer (DAP and NPK) purchased, and *Bought urea* is a dummy indicating whether the farmer purchased urea. *Solar Lamps* is a dummy variables equal to one if the farmer purchased any solar lamps. *Credit size* reports the size of the OAF loan. Standard errors are clustered at the farmer group level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table B7: Probability of Collecting Endline Data

	LPM			Odd ratios		
	Survey	Enroll 1st	Enroll 2nd	Survey	Enroll 1st	Enroll 2nd
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. KALRO						
Treated	0.019			1.325		
	(0.018)			(0.358)		
Mean Control	0.919			0.919		
Observations	833			833		
Panel B. IPA/PAD1-K						
Treated	0.014			1.086		
	(0.020)			(0.126)		
Mean Control	0.766			0.766		
Observations	1897			1897		
Panel C. IPA/PAD2-K						
Treated	-0.002			0.985		
	(0.012)			(0.077)		
Mean Control	0.820			0.820		
Observations	5890			5890		
Panel D. OAF1-K						
Treated	-0.012	-0.002	0.014	0.940	0.991	1.060
	(0.024)	(0.015)	(0.015)	(0.120)	(0.062)	(0.066)
Mean Control	0.750	0.602	0.397	0.750	0.602	0.397
Observations	1466	4884	4884	1466	4884	4884
Panel E. OAF2-K						
Treated		0.002	0.007		1.009	1.029
		(0.005)	(0.006)		(0.030)	(0.027)
Mean Control		0.761	0.558		0.761	0.558
Observations		32572	32572		32572	32572
Panel F. OAF3-R						
Treated		0.010	0.004		1.043	1.016
		(0.007)	(0.007)		(0.031)	(0.030)
Mean Control		0.647	0.475		0.647	0.475
Observations		86049	86049		86049	86049

Notes: The dependent variable in Panel A takes the value of one if the farmer completed the in person endline survey. In panels B and C the dependent variable indicates whether the farmer completed the phone-based endline survey. In panel D the dependent variable in columns (1) and (4) is a dummy variable indicating whether the farmer completed a phone-based survey (conducted with 30% of the original sample). In panels D-F, columns (2) and (5) the dependent variable indicates whether the farmer enrolled in the OAF program (i.e. placed an input order) in the season in which the program took place, while in columns (3) and (6) the dependent variable indicates whether they enrolled in the program in the following year. Columns (1)-(3) report marginal effects estimated using OLS, columns (4)-(6) report odds ratios estimated using Logit. * $p < .10$, ** $p < .05$, *** $p < .01$.

C Results by Experiment: Pooled Treatment Arms

Table C1: Awareness and Knowledge about Lime

	LPM				Logit (OR)			
	Heard Lime		Knows Lime Use		Heard Lime		Knows Lime Use	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. KALRO								
Treated	-0.007	-0.004	0.021	0.023	0.970	0.968	1.178	1.151
	(0.036)	(0.032)	(0.026)	(0.024)	(0.141)	(0.170)	(0.236)	(0.279)
Mean Control	0.58	0.58	0.14	0.14	0.58	0.58	0.14	0.14
Observations	773	773	773	773	773	773	773	773
Controls & FE	N	Y	N	Y	N	Y	N	Y
Panel B. IPA/PAD1-K								
Treated	0.037	0.037*	0.095***	0.093***	1.254*	1.338*	1.502***	1.751***
	(0.023)	(0.022)	(0.027)	(0.025)	(0.171)	(0.209)	(0.175)	(0.239)
Mean Control	0.78	0.78	0.33	0.33	0.78	0.77	0.33	0.33
Observations	1471	1471	1471	1471	1471	1435	1471	1471
Controls & FE	N	Y	N	Y	N	Y	N	Y
Panel C. IPA/PAD2-K								
Treated	0.055***	0.053***	0.099***	0.094***	1.509***	1.571***	1.488***	1.574***
	(0.013)	(0.012)	(0.017)	(0.016)	(0.133)	(0.156)	(0.099)	(0.119)
Mean Control	0.81	0.81	0.45	0.45	0.81	0.81	0.45	0.45
Observations	4822	4822	4822	4822	4822	4655	4822	4777
Controls & FE	N	Y	N	Y	N	Y	N	Y
Panel D. OAF1-K								
Treated	-0.002	0.001	0.096***	0.100***	0.990	0.998	1.515***	1.629***
	(0.026)	(0.025)	(0.031)	(0.030)	(0.159)	(0.174)	(0.204)	(0.237)
Mean Control	0.80	0.80	0.32	0.32	0.80	0.80	0.32	0.32
Observations	1087	1087	1087	1087	1087	1087	1087	1087
Controls & FE	N	Y	N	Y	N	Y	N	Y

Notes: This table reports the effect of each program on knowledge of agricultural lime. Columns (1) to (4) report marginal effects estimated using OLS, columns (5) to (8) report odds ratios, estimated using Logit. *Heard Lime* is a dummy variable reporting whether farmers had heard about agricultural lime before. *Knows Lime Use* is coded as one if the farmer mentions lime as a strategy to deal with or reduce soil acidity. Regressions in odd columns do not include any controls, regressions in even columns include controls and fixed effects. Robust standard errors shown in parenthesis. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C2: Followed Lime Recommendations

	LPM						Logit (OR)					
	Survey		Admin (all)		Admin (enrol)		Survey		Admin (all)		Admin (enrol)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. KALRO												
Treated	-0.004	-0.003	-0.012	-0.006			0.957	0.963	0.879	0.904		
	(0.021)	(0.020)	(0.022)	(0.022)			(0.233)	(0.281)	(0.210)	(0.239)		
Mean Control	0.10	0.10	0.11	0.11			0.10	0.14	0.11	0.12		
Observations	773	773	773	773			773	561	773	664		
Controls & FE	N	Y	N	Y			N	Y	N	Y		
Panel B. IPA/PAD1-K												
Treated	0.043*	0.038**	0.020	0.019			1.265*	1.539**	1.115	1.164		
	(0.024)	(0.017)	(0.021)	(0.017)			(0.165)	(0.308)	(0.126)	(0.173)		
Mean Control	0.22	0.22	0.24	0.24			0.22	0.22	0.24	0.25		
Observations	1471	1471	1897	1897			1471	1393	1897	1854		
Controls & FE	N	Y	N	Y			N	Y	N	Y		
Panel C. IPA/PAD2-K												
Treated	0.085***	0.076***	0.033**	0.030***			1.451***	1.663***	1.164**	1.379***		
	(0.016)	(0.013)	(0.014)	(0.009)			(0.103)	(0.150)	(0.076)	(0.145)		
Mean Control	0.31	0.31	0.30	0.30			0.31	0.31	0.30	0.28		
Observations	4822	4822	5890	5890			4822	4647	5890	5476		
Controls & FE	N	Y	N	Y			N	Y	N	Y		
Panel D. OAF1-K												
Treated	0.050**	0.041*	0.034***	0.030***	0.058***	0.054***	1.505**	1.564**	1.379***	1.395***	1.431***	1.495***
	(0.022)	(0.021)	(0.010)	(0.009)	(0.016)	(0.014)	(0.284)	(0.328)	(0.133)	(0.145)	(0.145)	(0.164)
Mean Control	0.12	0.12	0.10	0.10	0.17	0.17	0.12	0.12	0.10	0.10	0.17	0.17
Observations	1087	1087	4884	4884	2931	2931	1087	1087	4884	4884	2931	2931
Controls & FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Panel E. OAF2-K												
Treated			0.025***	0.025***	0.031***	0.031***			1.116***	1.153***	1.136***	1.201***
			(0.006)	(0.005)	(0.007)	(0.006)			(0.030)	(0.035)	(0.034)	(0.043)
Mean Control			0.32	0.32	0.42	0.42			0.32	0.32	0.42	0.42
Observations			32572	32572	24825	24825			32572	32572	24825	24623
Controls & FE			N	Y	N	Y			N	Y	N	Y
Panel F. OAF3-R												
Treated			0.006***	0.007***	0.008**	0.011***			1.169***	1.244***	1.155**	1.252***
			(0.002)	(0.002)	(0.003)	(0.003)			(0.071)	(0.068)	(0.070)	(0.069)
Mean Control			0.03	0.03	0.05	0.05			0.03	0.05	0.05	0.08
Observations			86049	86049	56303	56303			86049	57189	56303	39083
Controls & FE			N	Y	N	Y			N	Y	N	Y

Notes: This table reports the marginal effect of each program on whether farmers followed the lime recommendations. Columns (1)-(6) report marginal effects estimated using OLS. Columns (7)-(12) report odds ratios, estimated using Logit. Columns (1), (2), (7), and (8) report survey results. Column (3), (4), (9), and (10) show results for the administrative data (lime purchases or coupon redemption) for the entire sample of farmers participating in the experiment. Columns (5), (6), (11), and 12) show results for the administrative data for the subset of OAF farmers registered in the program. In panels A and D-F the dependent variable takes value one if the farmer used or acquired agricultural lime. In panels B and C, the dependent variable takes the value one if the farmer used lime in an area where it was recommended, or did not use lime in an area where it was not recommended. In panel A, columns (3), (4), (9), and (10) the results are measured through coupon redemption in the second season. Regressions in odd columns do not include any controls, regressions in even columns include controls and fixed effects. Robust standard errors shown in parenthesis. In panel F standard errors are clustered at the OAF group level. $^{*}p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$.

Table C3: Use of Recommended Fertilizers

	LPM						Logit (OR)					
	Survey		Admin (all)		Admin (enrol)		Survey		Admin (all)		Admin (enrol)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. KALRO												
Treated	-0.028	-0.029	0.028	0.030			0.839	0.824	1.122	1.151		
	(0.029)	(0.029)	(0.036)	(0.035)			(0.150)	(0.158)	(0.164)	(0.184)		
Mean Control	0.81	0.81	0.41	0.41			0.81	0.81	0.41	0.41		
Observations	773	773	773	773			773	773	773	773		
Controls & FE	N	Y	N	Y			N	Y	N	Y		
Panel B. IPA/PAD1-K												
Treated	0.010	0.010	0.012*	0.011			1.079	1.091	1.701	1.695		
	(0.020)	(0.020)	(0.007)	(0.007)			(0.164)	(0.180)	(0.590)	(0.617)		
Mean Control	0.15	0.15	0.02	0.02			0.15	0.17	0.02	0.03		
Observations	1471	1471	1897	1897			1471	1366	1897	1278		
Controls & FE	N	Y	N	Y			N	Y	N	Y		
Panel C. IPA/PAD2-K												
Treated	0.036***	0.034***	0.004	0.005			1.287***	1.296***	1.174	1.244		
	(0.012)	(0.013)	(0.005)	(0.005)			(0.116)	(0.124)	(0.228)	(0.256)		
Mean Control	0.16	0.16	0.02	0.02			0.16	0.16	0.02	0.04		
Observations	4822	4822	5890	5890			4822	4674	5890	3471		
Controls & FE	N	Y	N	Y			N	Y	N	Y		
Panel D. OAF2-K												
Treated			0.028***	0.030***	0.034***	0.033***			1.240***	1.345***	1.237***	1.348***
			(0.007)	(0.006)	(0.008)	(0.007)			(0.062)	(0.078)	(0.063)	(0.084)
Mean Control			0.14	0.14	0.19	0.19			0.14	0.14	0.19	0.19
Observations			32572	32572	24825	24825			32572	32572	24825	24825
Controls & FE			N	Y	N	Y			N	Y	N	Y

Notes: This table reports the effect of each program on use of chemical fertilizers. Columns (1) - (6) report marginal effects measured using OLS, columns (7) - (12) report odds ratios measured using Logit. In columns (1), (2), (7), and (8) the dependent variables are obtained self-reported survey data, while in columns (3) - (6) and (9) - (12) the dependent variables are measured through administrative data. In panel A, the dependent variable takes value one if the farmer used at least one type of recommended fertilizer, administrative data is obtained from coupon redemption in the second season. In panel B and C, the dependent variable in columns (1), (2), (7), and (8) indicates whether the farmer reported using urea, while the dependent variable in columns (3), (4), (9), and (10) indicates whether they used the electronic coupon to purchase urea. In panel D, the dependent variable indicates whether the farmer purchased "Extra CAN" from OAF. Since only a subset of treated farmers were recommended Extra CAN, here *Treated* indicates that the farmer was assigned to the "Lime+CAN" subtreatment. The regressions also include a dummy for the "Lime only" subtreatment. Columns (5), (6), (11), and (12) show results for the administrative data for the subset of OAF farmers registered in the program. Regressions in odd columns do not include any controls, regressions in even columns include controls and fixed effects. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C4: Lime Recommendations: Persistence & Fatigue

	LPM						Logit (OR)					
	Survey		Admin (all)		Admin (enrol)		Survey		Admin (all)		Admin (enrol)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. KALRO												
Treated (S_t)			-0.012	-0.006					0.879	0.904		
			(0.022)	(0.022)					(0.210)	(0.238)		
Mean Control			0.11	0.11					0.11	0.13		
Observations			773	773					773	664		
Controls & FE			N	Y					N	Y		
Panel B. IPA/PAD1-K												
Treated (S_t & S_{t+1})	0.058***	0.056***	0.004	0.006			1.474***	1.657***	1.039	1.121		
	(0.021)	(0.019)	(0.015)	(0.010)			(0.218)	(0.291)	(0.160)	(0.273)		
Mean Control	0.15	0.15	0.11	0.11			0.15	0.16	0.11	0.07		
Observations	1471	1471	1897	1897			1471	1404	1897	1531		
Controls & FE	N	Y	N	Y			N	Y	N	Y		
Panel C. IPA/PAD2-K												
Treated (S_t)	0.015	0.010*					1.089	1.098				
	(0.019)	(0.006)					(0.119)	(0.121)				
Mean Control	0.22	0.22					0.22	0.22				
Observations	2566	2566					2566	2566				
Controls & FE	N	Y					N	Y				
Panel D. OAF1-K												
Treated (S_t)			0.011	0.004	0.000	-0.011			1.163	1.088	1.002	0.912
			(0.015)	(0.014)	(0.022)	(0.021)			(0.234)	(0.227)	(0.206)	(0.194)
Treated (S_t & S_{t+1})			0.031**	0.023	0.030	0.018			1.455*	1.405*	1.286	1.229
			(0.015)	(0.015)	(0.023)	(0.022)			(0.287)	(0.289)	(0.260)	(0.257)
Mean Control			0.08	0.08	0.12	0.12			0.08	0.08	0.12	0.12
Observations			2931	2931	1986	1986			2931	2931	1986	1986
Controls & FE			N	Y	N	Y			N	Y	N	Y
Panel E. OAF2-K												
Treated (S_t)			0.004	0.005	0.006	0.007			1.060	1.090	1.072	1.084
			(0.004)	(0.003)	(0.006)	(0.005)			(0.065)	(0.064)	(0.067)	(0.065)
Treated (S_t & S_{t+1})			0.019***	0.018***	0.025***	0.023***			1.305***	1.339***	1.288***	1.318***
			(0.004)	(0.004)	(0.006)	(0.005)			(0.083)	(0.083)	(0.083)	(0.083)
Mean Control			0.07	0.07	0.10	0.10			0.07	0.09	0.10	0.11
Observations			51923	51923	36012	36012			51923	40628	36012	31468
Controls & FE			N	Y	N	Y			N	Y	N	Y

Notes: This table reports the effect of each program on whether farmers followed the lime recommendations for a second season. *Treated (S_t)* indicates that the farmer received SMS only in the first season, *Treated (S_{t+1})* indicates that the farmer received SMS only in the second season, *Treated (S_t & S_{t+1})* indicates that the farmer received SMS in both season. In panel E the comparison group is the set of farmers that received messages only in the second season. Columns (1)-(6) report marginal effects estimated using OLS. Columns (7)-(12) report odds ratios, estimated using Logit. Columns (1), (2), (7) and (8) report survey results. Columns (3), (4), (9), and (10) show results for the administrative data (lime purchases or coupon redemption) for the entire sample of farmers participating int the experiment. Columns (5), (6), (11), and (12) show results for the administrative data for the subset of OAF farmers registered in the program in the second season. In panels D, E, and F, the sample is restricted to the farmers registered for the program in the first season as the others were not eligible for receiving SMS messages in the second season. In panels A and D-F the dependent variable takes value one if the farmer used or acquired agricultural lime. In panels B and C, the dependent variable takes the value one if the farmer used lime in an area where it was recommended, or did not use lime in an area where it was not recommended. Regressions in odd columns do not include any controls, regressions in even columns include controls and fixed effects. The regression in panel C column (8) does not include fixed effects. Robust standard errors shown in parentheses. In panel F standard errors are clustered at the OAF group level. $^{*}p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$.

Table C5: Fertilizer Recommendations: Persistence

	LPM						Logit (OR)					
	Survey		Admin (all)		Admin (enrol)		Survey		Admin (all)		Admin (enrol)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. KALRO												
Treated (S_t)			0.028	0.030					1.122	1.151		
			(0.036)	(0.035)					(0.164)	(0.184)		
Mean Control			0.41	0.41					0.41	0.41		
Observations			773	773					773	773		
Controls & FE			N	Y					N	Y		
Panel B. IPA/PAD1-K												
Treated (S_t)	0.029	0.035*					1.217	1.327*				
	(0.021)	(0.020)					(0.177)	(0.214)				
Mean Control	0.17	0.17					0.17	0.18				
Observations	1471	1471					1471	1370				
Controls & FE	N	Y					N	Y				
Panel C. IPA/PAD2-K												
Treated (S_t)	-0.003	-0.004					0.970	0.965				
	(0.015)	(0.015)					(0.137)	(0.155)				
Mean Control	0.12	0.12					0.12	0.14				
Observations	2566	2566					2566	2111				
Controls & FE	N	Y					N	Y				
Panel D. OAF2-K												
Treated (S_t)			0.008	0.008	0.012	0.011			1.057	1.068	1.064	1.085
			(0.008)	(0.007)	(0.010)	(0.009)			(0.057)	(0.065)	(0.060)	(0.071)
Mean Control			0.18	0.18	0.24	0.24			0.18	0.18	0.24	0.19
Observations			24825	24825	18356	18356			24825	24825	18356	18356
Controls & FE			N	Y	N	Y			N	Y	N	Y

Notes: This table reports the effect of each program on whether farmers followed the fertilizer recommendations for a second season. *Treated (S_t)* indicates that the farmer received SMS only in the first season. Columns (1)-(6) report marginal effects estimated using OLS. Columns (7)-(12) report odds ratios, estimated using Logit. In panel A, the dependent variable takes value one if the farmer purchased at least one type recommended fertilizer. In panel B and C, the dependent variable indicates whether the farmer reported using urea. In panel D, the dependent variable indicates whether the farmer purchased "Extra CAN" from OAF. Since only a subset of treated farmers were recommended Extra CAN, here *Treated* indicates that the farmer was assigned to the "Lime+CAN" subtreatment. The regressions also include a dummy for the "Lime only" subtreatment. Columns (1), (2), (7), and (8) report survey results. Column (3), (4), (9), and (10) show results for the administrative data (fertilizer purchases or coupon redemption) for the entire sample of farmers participating in the experiment. In panel D the sample is restricted to the farmers registered for the program in the first season. Columns (5), (6), (11), and (12) show results for the administrative data for the subset of OAF farmers registered in the program in the second season. Regressions in odd columns do not include any controls, regressions in even columns include controls and fixed effects. Robust standard errors shown in parenthesis. $^{*}p < .10$, ** $p < .05$, *** $p < .01$.

Table C6: Use of All Recommended Inputs and Other Inputs

	Recommended Inputs (index)		Other Inputs (index)		Other Fertilizers (index)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. KALRO						
Treated	0.011	0.013	0.015	0.019		
	(0.034)	(0.033)	(0.044)	(0.039)		
Observations	773	773	773	773		
Controls & FE	N	Y	N	Y		
Panel B. IPA/PAD1-K						
Treated	0.055	0.056*			-0.081***	-0.079***
	(0.035)	(0.033)			(0.029)	(0.027)
Observations	1471	1471			1471	1471
Controls & FE	N	Y			N	Y
Panel C. IPA/PAD2-K						
Treated	0.065***	0.059***	-0.068***	-0.056**	-0.021	-0.022
	(0.020)	(0.017)	(0.025)	(0.024)	(0.020)	(0.019)
Observations	4822	4822	4822	4822	4822	4822
Controls & FE	N	Y	N	Y	N	Y
Panel D. OAF1-K						
Treated	0.102***	0.089***	0.030**	0.020		
	(0.029)	(0.027)	(0.014)	(0.013)		
Observations	4884	4884	4884	4884		
Controls & FE	N	Y	N	Y		
Panel E. OAF2-K						
Treated	0.075***	0.077***	0.000	-0.001		
	(0.015)	(0.013)	(0.009)	(0.009)		
Observations	13014	13014	13014	13014		
Controls & FE	N	Y	N	Y		
Panel F. OAF3-R						
Treated	0.029***	0.036***	0.009	0.003		
	(0.011)	(0.009)	(0.009)	(0.007)		
Observations	86049	86049	86049	86049		
Controls & FE	N	Y	N	Y		

Notes: This table presents results of indexes of recommended inputs (columns (1) and (2)), other inputs not mentioned by the SMS messages (columns (3) and (4)), and other fertilizers not recommended (columns (5) and (6)). Each index is composed of different variables depending on the project. For a full list of variables, see table A1. The coefficients are average effect sizes. Regressions in odd columns do not include any controls, regressions in even columns include controls and fixed effects (panel F, column (4) includes fixed effect at the OAF sector level instead of the site level to ensure standard errors can be computed). Robust standard errors in parentheses. In panel F standard errors are clustered at the OAF group level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C7: Quantities

	Kg Lime				Kg Fertilizer	
	Lime Rec.		Lime not Rec.			
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. KALRO						
Treated	-2.531	-1.787			1.481	1.463
	(3.757)	(3.738)			(0.918)	(0.931)
Mean Control	16.93	16.93			6.89	6.89
Observations	773	773			773	773
Controls & FE	N	Y			N	Y
Panel B. IPA/PAD1-K						
Treated	0.072	0.119	1.700	1.379	0.182	0.191
	(0.647)	(0.618)	(1.303)	(1.233)	(0.134)	(0.133)
Mean Control	2.85	2.85	3.32	3.32	0.24	0.24
Observations	1552	1552	345	345	1896	1896
Controls & FE	N	Y	N	Y	N	Y
Panel C. IPA/PAD2-K						
Treated	0.882*	0.966**	-1.620**	-1.495*	0.078	0.127
	(0.455)	(0.444)	(0.791)	(0.768)	(0.148)	(0.138)
Mean Control	3.52	3.52	3.56	3.56	0.55	0.55
Observations	4512	4512	1378	1378	5890	5890
Controls & FE	N	Y	N	Y	N	Y
Panel D. OAF1-K						
Treated	3.592***	3.207***				
	(0.821)	(0.794)				
Mean Control	5.82	5.82				
Observations	4884	4884				
Controls & FE	N	Y				
Panel E. OAF2-K						
Treated	2.258***	2.237***			1.451***	1.080***
	(0.482)	(0.449)			(0.492)	(0.407)
Mean Control	17.90	17.90			27.12	27.12
Observations	32572	32572			32572	32572
Controls & FE	N	Y			N	Y
Panel F. OAF3-R						
Treated	0.107	0.179				
	(0.141)	(0.119)				
Mean Control	1.73	1.73				
Observations	86049	86049				
Controls & FE	N	Y				

Notes: The table reports the effects of the programs on quantity of lime and fertilizer purchased, expressed in kgs. In panel A, columns (5) and (6), the dependent variable indicates the total quantity of fertilizer purchased (planting and top-dressing). In panel B and C, columns (1)-(2) and (3)-(4), the sample is divided based on whether lime was recommended in the farmer's area (Lime Rec) or not (Lime not Rec), while in columns (5) and (6) the dependent variable indicates the quantity of urea purchased using the electronic coupons. In panel E, columns (5) and (6) the dependent variable indicates the quantity of CAN purchased from OAF. Since only a subset of treated farmers were recommended Extra CAN, here *Treated* indicates that the farmer was assigned to the "Lime+CAN" subtreatment. The regressions also include a dummy for the "Lime only" subtreatment. All regressions include controls. Robust standard errors in parentheses. In panel E the standard errors are clustered at the OAF group level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C8: Spillovers

	LPM				Logit (OR)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. OAF1-K, Lime recommendations								
N treated	-0.001				0.988			
	(0.006)				(0.069)			
Sample								
Observations	Control				Control			
Controls & FE	1559				1453			
fe	Y				Y			
Panel B. OAF2-K, Lime recommendations								
N treated	0.006				1.038*			
	(0.004)				(0.024)			
Sample								
Observations	Control				Control			
Controls & FE	8142				7966			
fe	Y				Y			
Panel C. OAF2-K, Fertilizer recommendations								
N treated	0.004**				1.044**			
	(0.002)				(0.019)			
Sample	0.34				0.34			
Observations	Control & Lime Only				Control & Lime Only			
Controls & FE	27700				27700			
fe	Y				Y			
Panel D. OAF3-R, Lime recommendations								
N treated	-0.000		0.000		0.993		1.015	
	(0.001)		(0.000)		(0.022)		(0.012)	
Group Treated		0.004**		0.002**		1.045**		1.181**
		(0.002)		(0.001)		(0.071)		(0.096)
Mean Control		0.03		0.02		0.03		0.02
Sample	Partial Control	Partial & Full Control	All	All	Partial Control	Partial & Full Control	All	All
Has Phone	Yes	Yes	No	No	Yes	Yes	No	No
Observations	28520	48263	101891	101891	28520	48263	101891	101891
Controls & FE	Y	Y	Y	Y	Y	Y	Y	Y

Notes: This table reports spillover effects at the OAF group level. Columns (1)-(4) report marginal effects measured using OLS, columns (5)-(8) report odds ratios measured using Logit. In panel A, B, and D The dependent variable in column indicates whether farmers purchased lime from OAF. In panel B the dependent variable in column indicates whether farmers purchased the recommended fertilizer from OAF. *N treated* indicates the number of treated farmers in the OAF group, *Group treated* is a dummy equal to 1 if some farmers in the group were treated. The sample is restricted to farmers that were not assigned to receive messages or could not receive them because they did not have a valid phone number registered in the OAF database. All regressions include controls. Robust standard errors in parentheses, in panel C, columns (2)-(4) and (6)-(8) standard errors are clustered at the OAF group level. * $p < .10$, ** $p < .05$, *** $p < .01$.

D Pooled Regressions

Table D1: Pooled Regressions

	LPM		Logit (OR)	
	Lime	Fertilizer	Lime	Fertilizer
	(1)	(2)	(3)	(4)
Treated	0.013***	0.012***	1.143***	1.116***
	(0.002)	(0.004)	(0.020)	(0.040)
Mean Control	0.13	0.13	0.13	0.13
Observations	132125	41192	132125	41192

Notes: This table shows the effect of the programs on following lime and fertilizer recommendations, pooling data from all programs. Both dependent variables are measured using administrative data for the first season, except for KALRO, where administrative data for the second season is used. All regressions include program FEs. Columns (1)-(2) report marginal effects measured using OLS, columns (3)-(4) report odds ratios measured using Logit. Bootstrap standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table D2: Heterogeneity (Pooled Specifications)

	LPM						Logit (OR)					
	Female	Primary	Large Farm	Young	Used Input	Heard Input	Female	Primary	Large Farm	Young	Used Input	Heard Input
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Followed Lime Recommendations												
Treated	0.030***	0.021	0.013***	0.023***	0.007***	0.016	1.176***	1.143	1.137***	1.105***	1.142***	1.084
	(0.009)	(0.020)	(0.003)	(0.007)	(0.002)	(0.012)	(0.046)	(0.102)	(0.025)	(0.036)	(0.045)	(0.067)
[X]	0.053***	0.017	0.027***	-0.072***	0.076***	-0.029	1.315***	1.115	1.325***	0.704***	2.785***	0.840*
	(0.008)	(0.020)	(0.004)	(0.011)	(0.008)	(0.018)	(0.054)	(0.110)	(0.054)	(0.034)	(0.238)	(0.075)
[X] *Treated	-0.007	0.009	0.001	0.005	0.017	0.034	0.946	1.021	1.017	1.046	1.119	1.222*
	(0.011)	(0.025)	(0.004)	(0.013)	(0.010)	(0.025)	(0.046)	(0.114)	(0.046)	(0.058)	(0.123)	(0.148)
Mean Control	0.29	0.23	0.13	0.31	0.06	0.25	0.29	0.23	0.13	0.31	0.06	0.25
Observations	45029	9771	132125	40224	94669	8620	45029	9771	132125	40224	94669	8620
Panel B. Followed Fertilizer Recommendations												
Treated	0.098*	0.219	0.094**	0.098**	0.063	0.390	1.103	1.245	1.098**	1.103***	1.065	1.477
	(0.053)	(0.179)	(0.038)	(0.038)	(0.045)	(0.473)	(0.071)	(0.210)	(0.041)	(0.038)	(0.052)	(0.745)
[X]	0.205***	0.304*	-0.012	-0.292***	2.033***	0.039	1.227***	1.355**	0.988	0.747***	7.634***	1.040
	(0.062)	(0.178)	(0.069)	(0.068)	(0.075)	(0.372)	(0.087)	(0.207)	(0.056)	(0.043)	(0.548)	(0.347)
[X] *Treated	0.017	-0.068	0.064	0.043	0.183**	-0.300	1.017	0.934	1.066	1.043	1.200**	0.741
	(0.068)	(0.216)	(0.080)	(0.084)	(0.079)	(0.497)	(0.079)	(0.150)	(0.071)	(0.069)	(0.096)	(0.385)
Mean Control	0.13	0.08	0.13	0.13	0.13	0.38	0.13	0.08	0.13	0.13	0.13	0.38
Observations	40217	8620	41192	40224	41192	833	40217	8620	41192	40224	41192	833

Notes: This table shows results of heterogeneity analysis pooling data from all programs. The dependent variable is whether the farmer followed lime recommendations (panel A) or fertilizer recommendations (panel B) in the first season. Both dependent variables are measured using administrative data for the first season except for KALRO, where administrative data for the second season is used. We show results for gender, whether respondent completed primary school, whether the respondent's land is large (defined as above median use of inputs for the OAF samples and more than 1.5 acres of land for the other programs), whether the respondent was under 40 years old, whether the respondent had previously used the input, and whether the respondent had previous knowledge of the input. All regressions include program FEs. Columns (1)-(6) report marginal effects measured using OLS, columns (7)-(12) report odds ratios measured using Logit. Bootstrap standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

E Results by Experiment: By Treatment Arms

Table E1: Knowledge and Adoption by Treatment Arms

	LPM				Logit (OR)			
	Heard Lime	Knows Lime	Followed Lime Rec	Purchased Fertilizer	Heard Lime	Knows Lime	Followed Lime Rec	Purchased Fertilizer
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. IPA/PAD1-K								
General	0.038	0.066**	0.020	0.013	1.343	1.531***	1.188	1.822
	(0.025)	(0.028)	(0.020)	(0.009)	(0.246)	(0.240)	(0.206)	(0.733)
Specific	0.035	0.119***	0.017	0.010	1.333	1.986***	1.141	1.577
	(0.025)	(0.029)	(0.019)	(0.008)	(0.244)	(0.308)	(0.190)	(0.641)
Mean Control	0.78	0.33	0.24	0.02	0.77	0.33	0.25	0.03
Observations	1471	1471	1897	1897	1435	1471	1854	1278
p-value General=Specific	0.889	0.067	0.863	0.774	0.970	0.082	0.807	0.681
Controls & FE	Y	Y	Y	Y	Y	Y	Y	Y
Panel B. IPA/PAD2-K								
MS	0.042***	0.092***	0.029**	0.007	1.414***	1.533***	1.356**	1.331
	(0.015)	(0.020)	(0.013)	(0.006)	(0.168)	(0.137)	(0.179)	(0.322)
SMS + Call	0.070***	0.116***	0.029**	0.010*	1.881***	1.730***	1.318*	1.519*
	(0.014)	(0.019)	(0.014)	(0.006)	(0.231)	(0.156)	(0.187)	(0.368)
SMS + Call Offer	0.051***	0.089***	0.046***	-0.002	1.568***	1.520***	1.571***	0.899
	(0.015)	(0.020)	(0.013)	(0.005)	(0.192)	(0.139)	(0.209)	(0.239)
Mean Control	0.81	0.45	0.29	0.02	0.81	0.45	0.27	0.04
Observations	4822	4822	4822	5890	4638	4771	4387	3471
p-value SMS=SMS+Call	0.041	0.206	0.945	0.573	0.026	0.178	0.828	0.565
p-value SMS=SMS+Call Offer	0.553	0.882	0.222	0.134	0.416	0.927	0.237	0.118
p-value SMS+Call=SMS+Call Offer	0.155	0.162	0.212	0.040	0.167	0.159	0.183	0.039
Controls & FE	Y	Y	Y	Y	Y	Y	Y	Y
Panel C. OAF1-K								
Broad	0.006	0.091**	0.026**		1.019	1.573***	1.342**	
	(0.029)	(0.036)	(0.011)		(0.206)	(0.264)	(0.159)	
Detailed	-0.003	0.109***	0.034***		0.977	1.689***	1.450***	
	(0.030)	(0.035)	(0.011)		(0.199)	(0.279)	(0.171)	
Mean Control	0.80	0.32	0.10		0.80	0.32	0.10	
Observations	1087	1087	4884		1087	1087	4884	
p-value Broad=Detailed	0.764	0.631	0.501		0.836	0.661	0.485	
Controls & FE	Y	Y	Y		Y	Y	Y	
Panel D. OAF2-K								
Lime only			0.023***	0.009**			1.141***	1.102**
			(0.005)	(0.004)			(0.036)	(0.048)
Lime+CAN			0.032***	0.030***			1.202***	1.345***
			(0.008)	(0.006)			(0.052)	(0.078)
Mean Control			0.32	0.14			0.32	0.14
Observations			32572	32572			32572	32572
p-value Lime only=Lime+CAN			0.174	0.000			0.178	0.000
Controls & FE			Y	Y			Y	Y
Panel E. OAF3-R								
Full treatment			0.007***				1.238***	
			(0.002)				(0.074)	
Partial treatment: treated			0.007***				1.253***	
			(0.002)				(0.076)	
Mean Control			0.03				0.03	
Observations			86049				57189	
p-value Full treat=Partial treat			0.803				0.806	
Controls & FE			Y				Y	

Notes: The table shows the effect of each of the main treatments on knowledge of lime and probability to follow recommendations. Columns (1) - (4) report marginal effects estimated using OLS, columns (5) - (8) report odds ratios estimated using Logit. The dependent variable in column (1) is a dummy variable reporting whether farmers had heard about agricultural lime before. The dependent variable in column (2) is coded as one if the farmer mentions lime as a strategy to deal with or reduce soil acidity. The dependent variable in columns (3) indicates whether farmers followed lime recommendations, measured using administrative data. In panels A and B, it takes value one if the farmer used lime and lime was recommended or if farmer did not use lime and lime was not recommended, zero otherwise. In panels C-E takes value one if the farmer used lime, zero otherwise. The dependent variable in columns (4) indicates whether farmers followed fertilizer recommendations, measured using administrative data. All regressions include controls. Robust standard errors in parenthesis. In panel E the standard errors are clustered at the OAF group level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table E2: Message Framing

	LPM			Logit (OR)		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. OAF2-K, Lime						
Basic	0.017**			1.108**		
	(0.008)			(0.051)		
Yield Increase	0.034***	0.017*		1.217***	1.097*	
	(0.008)	(0.009)		(0.056)	(0.057)	
Experimentation (self)	0.027***	0.010		1.171***	1.056	
	(0.008)	(0.009)		(0.054)	(0.055)	
Experimentation (neighbors)	0.013*	-0.004		1.079	0.973	
	(0.008)	(0.009)		(0.050)	(0.051)	
Social Compasison	0.028***	0.010		1.174***	1.058	
	(0.008)	(0.009)		(0.054)	(0.056)	
Self-efficacy	0.028***	0.010		1.175***	1.059	
	(0.008)	(0.009)		(0.054)	(0.056)	
Family framed SMS			-0.016***			0.912***
			(0.005)			(0.028)
Mean Control	0.32			0.32		
F test p-value		0.23			0.23	
Observations	32572	24430	24430	32572	24430	24430
Includes Control Group	Y	N	N	Y	N	N
Controls & FE	Y	Y	Y	Y	Y	Y
Panel B. OAF2-K, Fertilizer						
Basic	0.025**			1.292**		
	(0.012)			(0.146)		
Yield Increase	0.019	-0.006		1.199	0.929	
	(0.012)	(0.016)		(0.145)	(0.148)	
Experimentation (self)	0.036***	0.011		1.414***	1.098	
	(0.012)	(0.016)		(0.162)	(0.168)	
Experimentation (neighbors)	0.029**	0.003		1.332**	1.033	
	(0.012)	(0.016)		(0.158)	(0.162)	
Social Compasison	0.043***	0.018		1.533***	1.192	
	(0.013)	(0.016)		(0.169)	(0.180)	
Self-efficacy	0.026**	0.000		1.296**	0.997	
	(0.012)	(0.016)		(0.152)	(0.156)	
Family framed SMS			-0.009			0.921
			(0.009)			(0.083)
Mean Control	0.14			0.14		
F test p-value		0.74			0.68	
Observations	32572	24430	24430	32572	24344	24344
Includes Control Group	Y	N	N	Y	N	N
Controls & FE	Y	Y	Y	Y	Y	Y
Panel C. OAF3-R, Lime						
General promotion	0.007***			1.262***		
	(0.002)			(0.096)		
Specific + yield impact	0.007***	-0.000		1.238***	0.985	
	(0.002)	(0.003)		(0.095)	(0.081)	
Self-diagnosis	0.009***	0.001		1.313***	1.040	
	(0.003)	(0.003)		(0.099)	(0.084)	
Soil test	0.007**	-0.001		1.226***	0.981	
	(0.003)	(0.003)		(0.097)	(0.081)	
How travertine works	0.006**	-0.001		1.216**	0.966	
	(0.003)	(0.003)		(0.093)	(0.080)	
Order immediately	0.007**	-0.001		1.256***	0.991	
	(0.003)	(0.003)		(0.099)	(0.082)	
Your cell is acidic + yield impact	0.006**	-0.001		1.201**	0.956	
	(0.003)	(0.003)		(0.093)	(0.079)	
SMS framed as gain			0.001			1.048
			(0.002)			(0.047)
Mean Control	0.03			0.05		
F test p-value		0.98			0.97	
Observations	86049	66306	66306	57189	42052	42052
Includes Control Group	Y	N	N	Y	N	N
Controls & FE	Y	Y	Y	Y	Y	Y

Notes: The table shows the effect of different framing of messages on lime purchases. Columns (1) - (3) report marginal effects estimated using OLS, columns (4) - (6) report odds ratios estimated using Logit. In panel A and C The dependent variable in column indicates whether farmers purchased lime from OAF. In panel B the dependent variable in column indicates whether farmers purchased the recommended fertilizer from OAF. All regressions include controls. Robust standard errors in parenthesis. In panel C the standard errors are clustered at the OAF group level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table E3: Number of Messages

	LPM			Logit (OR)		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. OAF2-K, Lime						
N Lime SMS	0.006***			1.035***		
	(0.001)			(0.008)		
N Lime SMS$\geq$ 1		-0.003			0.983	
		(0.012)			(0.068)	
N Lime SMS $\geq$ 2		0.025**	0.025**		1.159**	1.157**
		(0.012)	(0.012)		(0.083)	(0.083)
N Lime SMS $\geq$ 3		0.004	0.004		1.023	1.023
		(0.008)	(0.008)		(0.046)	(0.045)
N Lime SMS $\geq$ 4		0.004	0.004		1.023	1.022
		(0.008)	(0.008)		(0.045)	(0.045)
N Lime SMS = 5		-0.005	-0.005		0.973	0.974
		(0.008)	(0.008)		(0.044)	(0.043)
Mean Control	0.32	0.32		0.32	0.32	
Observations	32572	32572	24430	32572	32572	24430
Includes Control Group	Y	Y	N	Y	Y	N
Controls & FE	Y	Y	Y	Y	Y	Y
Panel B. OAF2-K, Fertilizer						
N Fert SMS	0.008***			1.074***		
	(0.002)			(0.015)		
N Fert SMS$\geq$ 1		0.021			1.279	
		(0.022)			(0.248)	
N Fert SMS $\geq$ 2		-0.001	-0.001		0.971	0.971
		(0.023)	(0.023)		(0.205)	(0.205)
N Fert SMS $\geq$ 3		0.021	0.023*		1.166	1.168
		(0.014)	(0.014)		(0.144)	(0.144)
N Fert SMS $\geq$ 4		-0.010	-0.010		0.912	0.909
		(0.014)	(0.014)		(0.112)	(0.112)
N Fert SMS = 5		-0.003	-0.002		1.010	1.011
		(0.014)	(0.014)		(0.122)	(0.122)
Mean Control	0.14	0.14		0.14	0.14	
Observations	32572	32572	24430	32572	32572	24430
Includes Control Group	Y	Y	N	Y	Y	N
Controls & FE	Y	Y	Y	Y	Y	Y
Panel C. OAF3-R, Lime						
N Lime SMS	0.002***			1.073***		
	(0.000)			(0.015)		
N Lime SMS$\geq$ 1		0.003			1.103	
		(0.002)			(0.074)	
N Lime SMS $\geq$ 2		0.005**	0.005**		1.153**	1.153**
		(0.002)	(0.002)		(0.070)	(0.070)
N Lime SMS $\geq$ 3		0.000	0.000		1.012	1.009
		(0.002)	(0.002)		(0.060)	(0.060)
N Lime SMS $\geq$ 4		0.001	0.001		1.022	1.025
		(0.002)	(0.002)		(0.061)	(0.061)
Mean Control	0.03	0.03		0.03	0.03	
Observations	86049	86049	66306	57189	57189	42052
Includes Control Group	Y	Y	N	Y	Y	N
Controls & FE	Y	Y	Y	Y	Y	Y

Notes: The table shows the effect number of messages on lime purchases. Columns (1) - (3) report marginal effects estimated using OLS, columns (4) - (6) report odds ratios estimated using Logit. In panel A and C The dependent variable in column indicates whether farmers purchased lime from OAF. In panel B the dependent variable in column indicates whether farmers purchased the recommended fertilizer from OAF. All regressions include controls. Robust standard errors in parenthesis. In panel C the standard errors are clustered at the OAF group level. * $p < .10$, ** $p < .05$, *** $p < .01$.

F Heterogeneity by Experiment

Table F1: Heterogeneous Effects in Following Lime Recommendations

	Logit (OR)					
[X]	Female	Primary School	Large Farm	Young	Used Input	Heard Input
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. KALRO						
Treated	0.615	1.334	1.117	0.811	0.926	0.670
	(0.255)	(0.558)	(0.396)	(0.333)	(0.249)	(0.225)
[X]	0.664	2.367**	1.362	0.962	0.713	0.504
	(0.260)	(1.002)	(0.586)	(0.355)	(0.571)	(0.211)
[X] *Treated	1.927	0.527	0.616	1.194	0.533	2.272
	(1.046)	(0.296)	(0.325)	(0.634)	(0.678)	(1.276)
Mean Control	0.12	0.12	0.12	0.12	0.12	0.12
Observations	664	664	664	664	664	664
Controls & FE	Y	Y	Y	Y	Y	Y
Panel B. IPA/PAD1-K						
Treated	1.222	0.993	1.202	1.343	1.179	1.062
	(0.226)	(0.249)	(0.241)	(0.263)	(0.190)	(0.165)
[X]	0.872	0.859	1.017	1.153	1.061	0.806
	(0.234)	(0.231)	(0.290)	(0.366)	(0.350)	(0.332)
[X] *Treated	0.870	1.277	0.928	0.684	0.914	1.642
	(0.270)	(0.403)	(0.283)	(0.208)	(0.377)	(0.735)
Mean Control	0.25	0.25	0.25	0.25	0.25	0.25
Observations	1854	1854	1854	1854	1854	1854
Controls & FE	Y	Y	Y	Y	Y	Y
Panel C. IPA/PAD2-K						
Treated	1.445***	1.355*	1.221*	1.292**	1.322***	1.226**
	(0.158)	(0.237)	(0.141)	(0.165)	(0.123)	(0.125)
[X]	1.316*	1.114	0.783	1.051	0.992	0.878
	(0.215)	(0.206)	(0.131)	(0.201)	(0.288)	(0.164)
[X] *Treated	0.768	0.966	1.236	1.046	0.982	1.344
	(0.143)	(0.197)	(0.225)	(0.186)	(0.319)	(0.280)
Mean Control	0.30	0.30	0.30	0.30	0.30	0.30
Observations	5732	5732	5732	5732	5732	5732
Controls & FE	Y	Y	Y	Y	Y	Y
Panel D. OAF1-K						
Treated	1.520**	1.478	1.391***	1.577		
	(0.271)	(0.440)	(0.164)	(0.443)		
[X]	1.070	0.745	0.766	0.725		
	(0.201)	(0.263)	(0.195)	(0.266)		
[X] *Treated	0.880	1.120	1.013	0.988		
	(0.193)	(0.459)	(0.253)	(0.425)		
Mean Control	0.10	0.11	0.10	0.11		
Observations	4812	1151	4884	1151		
Controls & FE	Y	Y	Y	Y		
Panel E. OAF2-K						
Treated	1.179***		1.161***	1.154***		
	(0.069)		(0.041)	(0.042)		
[X]	1.397***		0.749***	0.759***		
	(0.084)		(0.059)	(0.046)		
[X] *Treated	0.974		0.969	1.006		
	(0.067)		(0.071)	(0.070)		
Mean Control	0.33		0.32	0.33		
Observations	31597		32572	31604		
Controls & FE	Y		Y	Y		
Panel F. OAF3-R						
Treated			1.185**		1.230***	
			(0.079)		(0.072)	
[X]			1.297***		2.297***	
			(0.116)		(0.268)	
[X] *Treated			1.114		1.062	
			(0.106)		(0.139)	
Mean Control			0.05		0.05	
Observations			57189		57189	
Controls & FE			Y		Y	

Notes: This table shows results of heterogeneity analysis by sample. The dependent variable is whether the farmer followed the lime recommendations, measured using administrative data. In Panel A, results are measured through coupon redemption int the second season. We show results for gender, whether respondent completed primary school, whether the respondent's land is large (defined as above median use of inputs for the OAF samples and more than 1.5 acres of land for the other programs), whether the respondent was under 40 years old, whether the respondent had previously used the input, and whether the respondent had previous knowledge of the input. All regressions include controls. Effect sizes are reported in terms of odds ratios measured using Logit. In panel C, location fixed effects are removed to ensure convergence. Robust standard errors in parentheses. In panel F standard errors are clustered at the OAF group level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table F2: Heterogeneity in Following Fertilizer Recommendations

	Logit (OR)					
$[X]$	Female	Primary School	Large Farm	Young	Used Input	Heard Input
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. KALRO						
Treated	1.593*	1.140	1.077	1.090	1.110	1.409
	(0.440)	(0.273)	(0.229)	(0.260)	(0.280)	(0.658)
$[X]$	1.529	1.292	0.885	0.941	1.408	1.106
	(0.397)	(0.310)	(0.265)	(0.224)	(0.374)	(0.435)
$[X]$ *Treated	0.611	1.004	1.177	1.102	1.057	0.797
	(0.208)	(0.329)	(0.389)	(0.360)	(0.346)	(0.398)
Mean Control	0.41	0.41	0.41	0.41	0.41	0.41
Observations	773	773	773	773	773	773
Controls & FE	Y	Y	Y	Y	Y	Y
Panel B. IPA/PAD1-K						
Treated	2.124	1.394	2.036	1.629	1.548	
	(0.992)	(0.843)	(1.001)	(0.702)	(0.650)	
$[X]$	1.345	1.042	1.690	0.494	2.026	
	(0.905)	(0.710)	(1.120)	(0.425)	(1.519)	
$[X]$ *Treated	0.535	1.338	0.632	1.209	1.402	
	(0.410)	(1.019)	(0.473)	(1.016)	(1.180)	
Mean Control	0.03	0.03	0.03	0.03	0.03	
Observations	1278	1278	1278	1278	1278	
Controls & FE	Y	Y	Y	Y	Y	
Panel C. IPA/PAD2-K						
Treated	1.011	1.512	1.443	1.145	1.454	
	(0.237)	(0.644)	(0.390)	(0.316)	(0.341)	
$[X]$	0.695	1.641	1.389	0.919	2.796***	
	(0.275)	(0.715)	(0.507)	(0.391)	(1.095)	
$[X]$ *Treated	1.737	0.755	0.655	1.103	0.487	
	(0.777)	(0.364)	(0.260)	(0.442)	(0.216)	
Mean Control	0.03	0.03	0.03	0.03	0.03	
Observations	4024	4024	4024	4024	4024	
Controls & FE	Y	Y	Y	Y	Y	
Panel D. OAF2-K						
Treated	1.391***		1.356***	1.332***	1.324***	
	(0.135)		(0.087)	(0.088)	(0.088)	
$[X]$	1.223***		0.803***	0.865***	6.305***	
	(0.055)		(0.062)	(0.040)	(0.285)	
$[X]$ *Treated	0.973		0.963	1.087	1.055	
	(0.105)		(0.109)	(0.119)	(0.115)	
Mean Control	0.15		0.14	0.15	0.14	
Observations	31597		32572	31604	32572	
Controls & FE	Y		Y	Y	Y	

Notes: This table shows results of heterogeneity analysis by sample. The dependent variable is whether the farmer followed the fertilizer recommendations, measured using administrative data. In Panel A, results are measured through coupon redemption int the second season. We show results for gender, whether respondent completed primary school, whether the respondent's land is large (defined as above median use of inputs for the OAF samples and more than 1.5 acres of land for the other programs), whether the respondent was under 40 years old, whether the respondent had previously used the input, and whether the respondent had previous knowledge of the input. All regressions include controls. In panel C, location fixed effects are removed to ensure convergence. Effect sizes are reported in terms of odds ratios measured using Logit. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

G Additional Meta-analyses

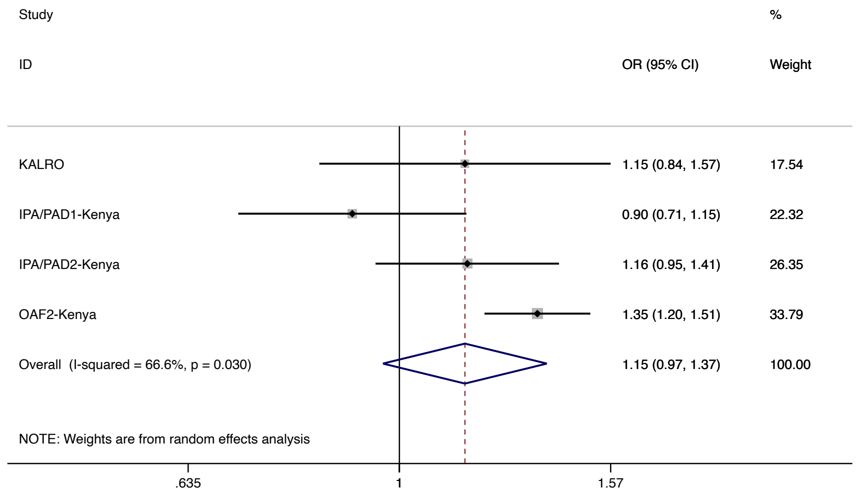
G.1 Alternative Assumptions

Table G1: Additional Meta-analysis Results

			Effects			Heterogeneity				
Row #	Outcome	N	Effect	95% CI		Q stat (p-value)	I^2	I^2 - 95% CI		τ^2
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Linear Probability Model										
	LPM									
1	Heard Lime	4	0.03	0.00	0.06	0.14	44.58	0.00	81.50	0.00
2	Knowledge Acidity	4	0.08	0.04	0.11	0.07	57.82	0.00	85.99	0.00
3	Lime Rec.	6	0.02	0.01	0.03	0.00	75.93	45.87	89.29	0.00
4	Fert Rec.	4	0.02	0.00	0.03	0.01	74.66	29.51	90.89	0.00
5	Peristence Lime	4	0.01	0.00	0.01	0.84	0.00	0.00	84.69	0.00
6	Fatigue Lime	3	0.02	0.01	0.02	0.50	0.00	0.00	89.60	0.00
7	Peristence Fert.	4	0.01	-0.00	0.02	0.42	0.00	0.00	84.69	0.00
	Standard Deviations									
8	Recomm Inputs (Index)	6	0.06	0.03	0.08	0.07	51.53	0.00	80.69	0.00
9	Other Inputs (Index)	5	0.00	-0.02	0.02	0.09	50.49	0.00	81.84	0.00
	Kg									
10	Kg Lime	5	1.07	-0.35	2.50	0.00	89.77	78.97	95.03	2.14
11	Kg Fertilizer	4	0.33	-0.01	0.68	0.08	55.60	0.00	85.30	0.06
Panel B. Odds Ratios, Hartung-Knapp-Sidik-Jonkman										
1	Heard Lime	4	1.23	0.84	1.80	0.03	65.87	0.00	88.37	0.04
2	Knowledge Acidity	4	1.57	1.27	1.95	0.51	0.00	0.00	84.69	0.01
3	Lime Rec.	6	1.23	1.10	1.38	0.21	29.37	0.00	71.10	0.01
4	Fertilizer Rec.	4	1.31	1.12	1.52	0.71	0.00	0.00	84.69	0.01
5	All Recommended Inputs	6	1.20	1.11	1.29	0.61	0.00	0.00	74.62	0.01
6	Other Inputs	5	1.00	0.85	1.17	0.06	55.21	0.00	83.46	0.01
7	Persistence Lime	4	1.08	1.01	1.16	0.92	0.00	0.00	84.69	0.00
8	Fatigue Lime	3	1.33	1.13	1.55	0.75	0.00	0.00	89.60	0.01
9	Persistence Fert.	4	1.10	0.92	1.31	0.51	0.00	0.00	84.69	0.01
Panel C. Odds Ratios, Empirical Bayes										
1	Heard Lime	4	1.23	0.97	1.56	0.03	65.87	0.00	88.37	0.03
2	Knowledge Acidity	4	1.58	1.41	1.78	0.51	0.00	0.00	84.69	0.00
3	Lime Rec.	6	1.22	1.13	1.31	0.21	29.37	0.00	71.10	0.00
4	Fertilizer Rec.	4	1.32	1.19	1.47	0.71	0.00	0.00	84.69	0.00
5	All Recommended Inputs	6	1.20	1.14	1.26	0.61	0.00	0.00	74.62	0.00
6	Other Inputs	5	1.00	0.89	1.12	0.06	55.21	0.00	83.46	0.01
7	Persistence Lime	4	1.0	0.98	1.19	0.92	0.00	0.00	84.69	0.00
8	Fatigue Lime	3	1.33	1.19	1.49	0.75	0.00	0.00	89.60	0.00
9	Persistence Fert.	4	1.09	0.99	1.209	0.52	0.00	0.00	84.69	0.00

Notes: Meta-analysis results for each outcome reported in the rows. Column (2)-(5) reports results from a random-effects model; Column (6)-(9) reports heterogeneity results. The coefficient represents the estimated summarized effects across studies. Panel A, rows 1-7 reports LPM results, rows 8-9 report results measured in standard deviations. Panel B and Panel C reports results in terms of Odd Ratios.

Figure G1: Effects on Any Mentioned Fertilizer (Administrative and Survey)



Notes: The figure plots the meta-analysis results for following fertilizer recommendations using administrative data. The effects are estimated using a random-effects meta-analysis model. Results are reported in odds ratios. The horizontal lines denote 95% confidence intervals. The results are measured using administrative data. The KALRO results are measured using coupon redemption in the second season. The dependent variable for IPA/PAD1-Kenya is a dummy equal to one if either DAP or urea were purchased. The dependent variable for IPA/PAD2-Kenya is a dummy equal to one if DAP, urea, or CAN were purchased.

G.2 Bayesian Meta-analysis

As a complementary exercise, we re-examine our results using a Bayesian hierarchical framework (Rubin, 1981; Gelman et al., 2013). The main difference with the random-effects model underlying the frequentist meta-analysis presented in the main paper is that in this case we define (weakly informative) prior distributions for the between-study heterogeneity τ^2 and true effect size μ. An additional advantage of this approach is that uncertainty of the estimate of τ^2 can be directly modeled and a posterior distribution for μ can be obtained. For a discussion of Bayesian hierarchical models with applications to economics, see Meager (2019) and Vivalt (2016). The analysis was implemented using R's baggr's Rubin (1981) model with default priors on the hyper-standard-deviation and hypermean (zero centered and scaled to data) (Wiecek and Meager, 2022).[46]

Figures G2 show forest plots for partially pooled models. In this case, while each project is assumed to have a different chance of success, the data for all the projects informs the estimates for each project. In other words, the bayesian estimation is a weighted average of each project and the average effect across all programs. The idea is that the model 'pools' power across projects, since projects can provide valuable information about one another. Table G2 summarizes the main results.

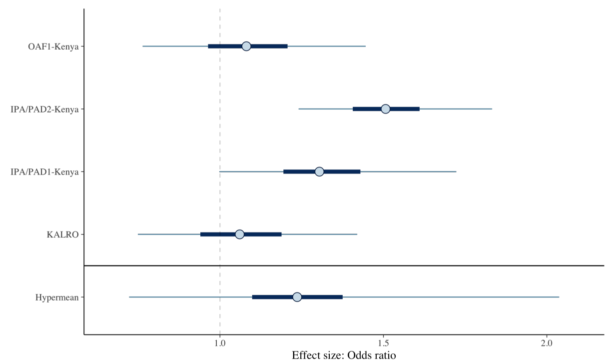
Table G2: Bayesian Hierarchical Models

				Effects		Heterogeneity		
Row #	Outcome	N	Effect	95% CI		I^2	I^2 - 95% CI	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Odds Ratios								
1	Heard Lime	4	1.27	0.72	2.04	70.3	5.97	98.7
2	Knowledge Acidity	4	1.57	1.11	2.07	39.6	0.09	96.3
3	Lime Rec.	6	1.22	1.08	1.39	28.9	0.08	83.9
4	Fertilizer Rec.	4	1.33	0.94	1.81	32.7	0.07	93.7
5	All Recommended Inputs	6	1.20	1.09	1.32	24.5	0.03	84.0
6	Other Inputs	5	1.01	0.83	1.19	56.5	0.36	97.5
7	Persistence Lime	4	1.07	0.83	1.32	29.2	0.04	90.6
8	Fatigue Lime	4	1.11	0.84	1.44	42.2	0.11	96.7
9	Persistence Fert.	3	1.34	0.84	2.11	46.0	0.13	96.5

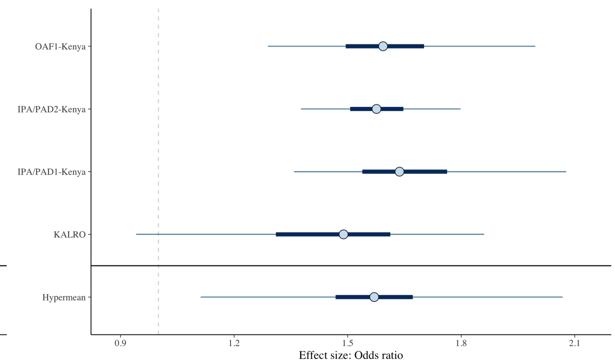
Notes: Meta-analysis results for each outcome reported in the rows. Columns (2)-(4) reports effects (in odds ratios); column (5)-(7) reports heterogeneity results.

[46]Qualitatively similar results are obtained under different priors. For instance, normal (0,10) priors on the hyper-standard-deviation for the first steps.

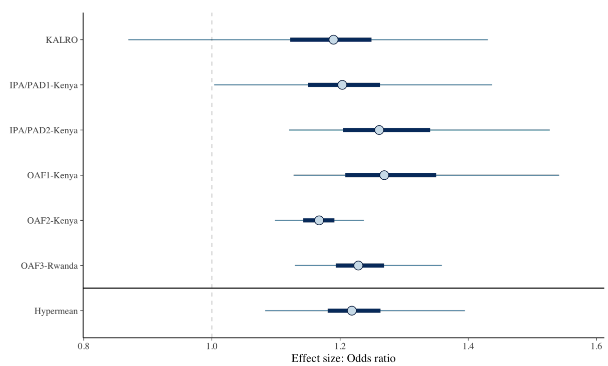
Figure G2: Bayesian Meta-analysis Effects



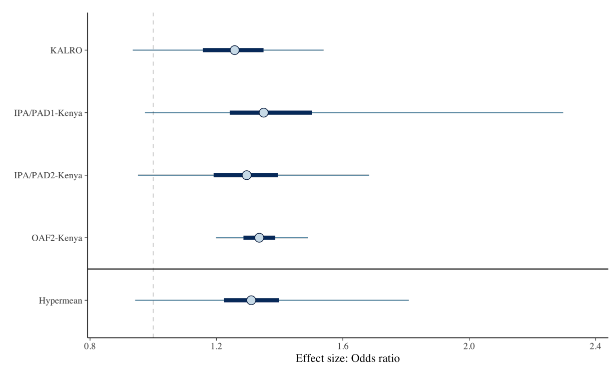
(a) Awareness (Lime)



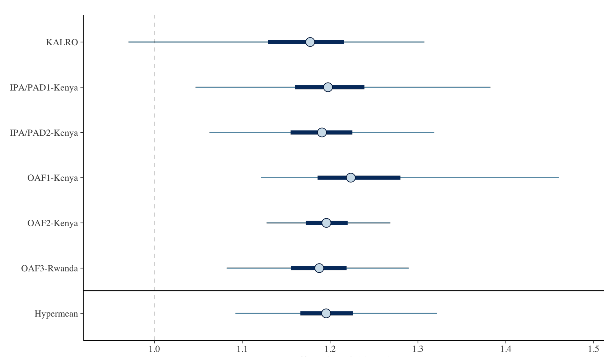
(b) Knowledge (Lime)



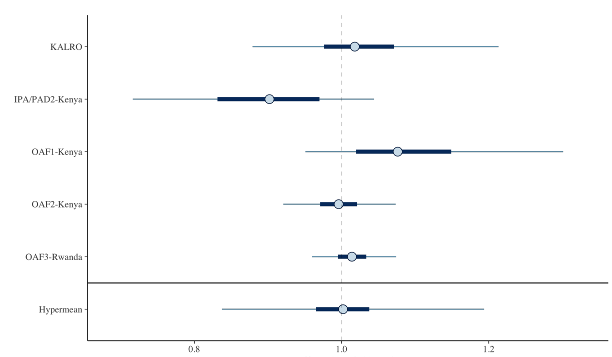
(c) Followed Recommendations (Lime)



(d) Followed Recommendations (Fertilizer)



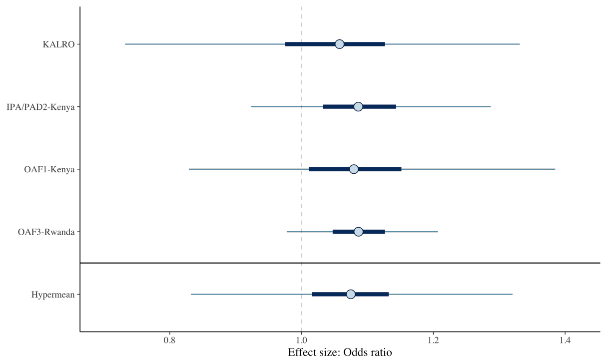
(e) Recommended Inputs



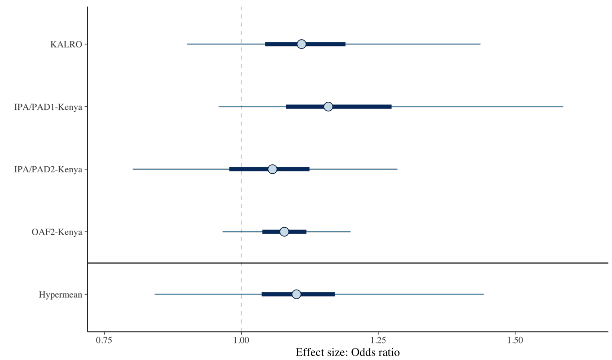
(f) Other Inputs

Notes: The figure plots the meta-analysis results for specific outcomes. The effects are estimated using bayesian hierarchical models. Results are reported in odds ratios. The horizontal lines denote 95% confidence intervals.

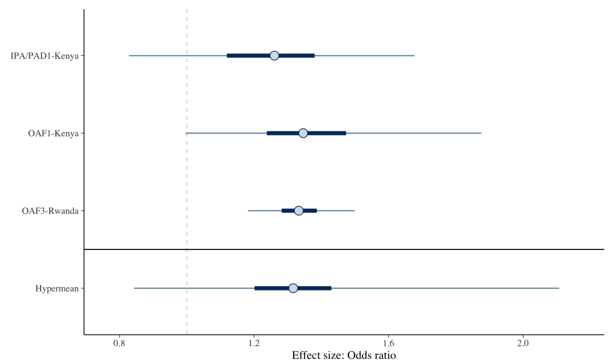
Figure G3: Bayesian Meta-analysis Effects (cont'd)



(a) Persistence (Lime)



(b) Persistence (Fertilizers)



(c) Fatigue (Lime)

Notes: The figure plots the meta-analysis results for specific outcomes. The effects are estimated using bayesian hierarchical models. Results are reported in odds ratios. The horizontal lines denote 95% confidence intervals.

H Additional Program and Experiment Details

H.1 KALRO's Program

The Kenya Agriculture and Livestock Research Organization (KALRO) is a public agency with the mandate to promote agricultural research and dissemination in Kenya. In 2014 and 2015, KALRO's Kakamega office implemented two extension programs aimed at encouraging smallholder farmers to adopt inputs and management practices that could address some of the regional soil deficiencies. These programs reflected their goal of reaching a large number of farmers at a lower cost than that of in-person individual farm visits.[47]

KALRO's SMS program consisted of sending 20 different agriculture-related text messages to maize farmers' mobile phones. The content of the messages was developed by the Ministry of Agriculture, Livestock and Fisheries and the delivery was managed by KALRO.[48] KALRO's e-extension program consisted in 20 SMS messages send in the period corresponding to the 2015 short rains season: June-November 2015. The first set of messages were in English. Mid-intervention the messages were switched to Swahili.[49] We list all messages sent by KALRO below:

- We at KALRO- Kakamega shall be sending you 20 SMS tips on how to increase your maize and legume (beans, groundnuts, soybeans) yield
- Keep all the records of your farming activities including inputs and outputs to help you know whether your farming is profitable
- Test your soil after every 4 years. Enquiries: KALRO Tel:[phone] or Soil Cares Ltd: [phone]
- If soil is acidic (pH less than 5.5), apply recommended rate of agricultural lime at least 30 days before planting. Enquiries: Tel.[phone]
- Construct raised bands and trenches to control soil erosion, reduce nutrient loss and keep rain water in the soil
- Add and/or leave all organic matter (manure, crop/weed residues and compost) to your field. Do not burn your fields. Burning destroys useful micro-organisms.
- Prepare land early, at least one plough and one harrow, ready for planting before onset of rains
- Plant before or at the onset of rains. Plant on well drained, fertile soils
- Use certified maize and legume seed recommended for your area, bought from an approved agro-dealer. Use 10 kg maize seed and 40kg of legume seed per acre. Enquiries: [phone]
- Maize and legumes planted in rows are easier to weed & apply fertilizer. You may plant maize alone/pure or together with legumes as follows:
- For pure maize make rows 2.5 feet (75cm) apart and holes 1 foot (30cm) apart along the row. Place 2 and 1 maize seeds in alternate holes.
- For maize and legume intercrop, plant maize as for pure stand and one row of legume (beans, soybean or groundnut) between two maize rows at spacing of 10cm from one hole to another.
- For better maize and legume harvests, inoculate legumes, rotate or intercrop, use fertilizer and manage your crop and soils appropriately.
- Use fertilizer to increases yields. Apply 1 heaped Fanta top of NPK or DAP in each hole for maize, cover with little soil, add seed and cover seed with soil. Fertilizer MUST not touch the seed
- Weeds compete with your crops for nutrients and so reduce yields. Keep fields free of weeds and pests. Thin maize seedlings to 1 plant per hole as you weed.
- Topdress your maize with a level Fanta bottle top of CAN or Mavuno top dress fertilizers 6 weeks after planting. Apply around each plant-5cm away and cover with soil. Apply when soil is moist.

[47]KALRO experimented with two approaches. First, with farmer field days (FFD), one-day events in which a large number of farmers can observe demonstration plots and receive information from extension agents. The second approach consisted of delivering agricultural messages to farmers via SMS. This paper focuses on the results from the second approach, but we discuss further details of the impacts of FFDs in Fabregas et al. (2017a).

[48]Since 2014 the Ministry of Agriculture has announced plans to roll out an e-extension system to reach over 7 million farmers, by providing phone-based support to extension workers who would then advise farmers. The version of the program that was evaluated was a pilot program that tried to deliver information directly to farmers. In July 2018, the Kenyan Ministry of Agriculture and Irrigation, in partnership with PAD and Safaricom, launched an SMS service (MoA-Info) aimed at providing agricultural advice to farmers across the country.

[49]While 75% of farmers report speaking English at baseline, there is a risk that some farmers might have not understood the initial messages. We do not find heterogeneous treatment effects by language spoken.

- Harvest as soon as the crops are mature. For maize look for the black eye; for legumes when 90-100% of pods are brown. In late harvests, termites, rodents, insects, diseases & birds reduce yield.
- Remove husk from maize cobs in the field to avoid transporting weevils from the field to the store. The husks will improve the organic matter in the soil.
- Dry your harvest in open sun, but protect it from rain. Thresh/shell and re-dry to moisture content of 11-12%.
- Store your harvest well in silos and helmetic bags. You may also use superactellic during storage – the insects will not touch the grain & is safe.
- Obtain information on favorable market prices before you sell your harvest

To recruit farmers into the program the IPA evaluation team conducted a census of farmers in the Ugenya and Mumias sub-counties using a walking rules to visit a representative sample of households. Farmers who owned a mobile phone, had grown maize or legumes during the previous year, and were in charge of farming activities in the household were then invited to participate in the project. Enumerators completed a total of 1,330 census surveys, and approximately 94% of those recruited during census activities met the selection criteria.

In September 2014, farmers completed an in-person baseline survey and were then randomized into the SMS treatment (415 farmers) or a comparison group (417 farmers). The text-message service was implemented between July 2015 and November 2015, in the period corresponding to the short rains season. An in-person endline survey, asking information about input use and knowledge, was completed with 92% of the baseline sample by January 2016. At the end of the endline survey, all farmers received two (paper) discount coupons redeemable at selected agricultural supply dealers in their nearest market center. The coupons were devised as a way to collect information on input choices and reduce concerns about enumerator demand effects since purchasing decisions would be made at a later time when farmers were not directly observed by any member of the research or KALRO teams. The first discount coupon was redeemable for a 50% discount for agricultural lime. The second coupon was redeemable for a 50% discount for any chemical fertilizer of their choice (NPK, DAP, CAN, urea or Mavuno). Both coupons had an upper limit discount of approximately \$10 USD. Coupon redemption was possible up to the start of the subsequent 2016 long rain agricultural season (March 2016). Participating agricultural supply dealers were instructed (and incentivized through a small payment) to keep clear records on input choices and quantities purchased by farmers who redeemed coupons. Incentives for shopkeepers were paid on the basis of having both the physical coupon and a record of the purchase in their logbooks.

H.2 IPA & PAD's Programs

PAD supported two agricultural extension research projects in western Kenya that were implemented and evaluated by IPA.

H.2.1 Program 1 (IPA/PAD1-K)

Throughout the 2016 short rain agricultural season, IPA, with support from PAD, sent selected farmers text messages with information about agricultural inputs (including lime and chemical fertilizers) as well as other general agronomic recommendations on maize farming. Farmers who participated in this program were recruited through administrative farmer records of a large agribusiness in the region (MSC farmers) and from records of individuals who had participated in IPA's activities previously (IPA farmers).[50] In July 2016, a random sample of farmers from both databases were contacted over the phone to invite them to participate in the study and complete a short phone-based baseline survey to determine eligibility. Farmers who were planning to plant maize in the 2016 short rains season, had a farm located within the intervention area, and expressed interested in receiving agricultural information over their phone were invited to participate.From 2,255 interviewed respondents, 2,131 consented to participate in the baseline and 1,897 (89%) met the criteria for selection.

This final sample was randomized into three groups: receiving the general messages ("General"), receiving specific messages for their area ("Specific"), and a control group. In addition, during the following agricultural season (long rains 2017) both treatment groups received five additional messages promoting the use of agricultural lime.

In order to construct customized recommendations for the specific messages, farmers were linked to a local landmark that could then be matched with soil data. This is a context in which there are no addresses and a lot of variation on how village names are reported. Therefore we used primary schools as landmarks for a given area. IPA farmers were matched to their closest primary school and provided recommendation based on median soil characteristics (exchangeable acidity and phosphorous) obtained from soil tests performed in the 2 km area around the school. The soil data were collected for previous projects by IPA (Fabregas et al., 2017b) and analyzed by the Kenya Agricultural Research Institute (KARI) using wet chemistry in 2011 and 2014. MSC farmers were matched to their "field", a set of plots cultivated by multiple farmers and aggregated by the company for organizing their activity, including soil testing. The recommendations provided to them were based on median soil characteristics

[50] The Mumias Sugar Company ran a contract farming model with sugar cane farmers in the region up to 2015. The vast majority of farmers plants maize in addition to many sugar cane so the company supported delivery of maize extension messages. The farmers who appeared in the IPA database were mainly recruited through large school meetings, as discussed in Duflo et al. (2018). This group accounted for about 47% of the final sample.

(pH and phosphorus) of the sample collected from that field and analyzed by MSC in the period 2009-2016.[51]

Among farmers receiving these messages with specific information, those who lived in areas that had median pH of more than 5.5 did not receive message about lime (18% of the sample). Both groups of farmers also received messages about planting (DAP) and at top-dressing (urea) fertilizers. Farmers received between 24 and 28 messages. Messages were sent either in English or in Swahili, depending on farmers' preferences indicated during the baseline phone survey. We report all the messages below: [G] indicates that the message was received by the General treatment group, and [S] denotes it was received by the Specific treatment group.

- [G/S]: Welcome to PAD's SMS information service. We will give you tips on agricultural inputs to apply on 1/8 of an acre so you can experiment during this short rains season. Receiving SMS messages is free.
- [G]: High soil acidity levels reduce nutrients available to plants, such as phosphorus, which causes symptoms of stunted growth and purple colouration of maize.
- [S]: Previous soil tests of shambas around [landmark] showed [degree] soil acidity levels. High acidity levels reduce nutrients available to plants, such as phosphorus, which causes symptoms of stunted growth and purple colouration of maize.
- [G]: Lime reduces soil acidity and makes nutrients such as phosphorus available for your maize.
- [S]: Based on soil tests of shambas around [landmark], we recommend you buy [quantity] kg of lime, [quantity] kg of DAP, and 6 kg of urea for microdosing 1/8 acre of your maize. Lime reduces soil acidity and makes phosphorus available for your maize.
- [S]: We would like you to try our recommendations in 1/8 of an acre. To measure 1/8 of an acre, walk around your farm and draw a square with each side 33 steps long. Walk normally, don't make long strides. If you land is a rectangle, the sum of 2 sides should measure in total 66 steps. Start from a corner, walk along the short side, count your steps until you reach the end. Turn around and keep walking along the long side until you finish counting 66 steps.
- [S]: When planting this season try adding a layer of lime [quantity] bottletop, then cover with soil and add a second layer of DAP ([quantity] bottletop) per hole on 1/8 acre to correct soil acidity and make more nutrients available for your plants. Apply 1 bottletop of urea per hole at top dressing.
- [G]: Use a ruler or measured rope to plant maize in rows using correct spacing of 75 cm x 25 cm. This offers maximum yield while limiting competition for nutrients, light and water.
- [S]: Use a ruler or measured rope to plant maize in rows using correct spacing of 75 cm x 25 cm. This offers maximum yield while limiting competition for nutrients, light, and water. You should be able to fit 2580 planting holes in 1/8 of an acre. Use sisal twine to encircle this area so you can compare the results at harvest.
- [S]: Have you bought lime and DAP yet? If not, buy a total of [quantity] kg of lime and use with [quantity] kg DAP for microdosing on 1/8 of your acre. DAP is the most cost efficient source of phosphorous. When lime is combined with DAP, it reduces soil acidity and makes nutrients available for your maize.
- [G]: Calcium lime is safer for your health and the plant. This lime could be either brown or grey.
- [S]: [agrovet] will be stocked with lime (calcium lime) and DAP during this short rain season. This lime is brown and it is safer for your health and the plant. It is also heavier than the white lime so you only need to apply [quantity] bottletop per plant. The price of lime today is Ksh 7 per kg. The price of DAP today is Ksh [price] per kg.
- [G/S]: Plant maize seed when there is enough moisture after 2-3 rains, to enable absorption of water by seed and fertilizer. Delayed planting leads to reduced yields. To stop receiving these SMS messages reply "STOP".
- [G/S]: Plant two maize seeds per hole to ensure one survives. Do not use broken or damaged seeds because they will not germinate. Use certified seeds, they grow faster and are high yielding.
- [G]: Are you ready to plant your maize? We recommend you apply both lime and fertilizer in micro-doses at planting. 5 weeks later we recommend you apply top dressing fertilizer in micro-doses
- [S]: Do you know the 5 Golden Rules for successful micro-dosing? Based on soil tests performed around [landmark], we recommend you to: Apply [quantity] bottletop of lime and cover with soil and then add [quantity] bottletop of DAP. Cover with 2 inches of soil. Use 2 seeds per planting hole.Cover the seeds with 2 inches of loose soil. Apply 1 bottletop of urea as top dressing fertilizer 5 weeks later when the plant is knee high.
- [G/S]: Remember, lime should only be used during planting and not at top dressing. Lime is not a fertilizer and could burn the plant if applied at top dressing.
- [G/S]: At planting, if you are applying lime in micro-doses, remember to cover it with soil before applying fertilizer and planting seeds. Lime should not be in direct contact with the seeds as it may burn them. When you apply lime, wear protective clothing such as long sleeves and gloves. Cover your mouth and nose with a scarf and wear goggles.
- [G/S]: Gap your maize immediately after emergence. Gapping is done by re-planting maize seeds in places that have not germinated. This gives you optimum plant population that leads to optimum yields.
- [G/S]: During first weeding, thin to one maize plant per hole. You should remove striga immediately to reduce competition for nutrients and water, and to prevent stunted growth!
- [G]: Have you already planted your maize this season? If not, we recommend applying lime at planting. Lime reduces soil acidity and makes nutrients such as phosphorus available for your maize.
- [S]: Have you already planted your maize this season? If not, we recommend applying lime at planting. We recommend you apply [quantity] bottletop per planting hole. Buy [quantity] kg of lime to experiment on 1/8 of an acre. Lime reduces soil acidity and makes nutrients such as phosphorus available for your maize.
- [G]: If you applied lime on your maize at planting, we recommend using urea at top dressing because it is a less expensive source of nitrogen.
- [S]: If you applied lime on your maize at planting, we recommend using urea for top dressing because it is a less expensive source of nitrogen. Buy 6 kg of urea for use on 1/8 of an acre.
- [S]: [agrovet] will be stocked with urea during this short rain season. The price of urea is Ksh [agrovet] per kg.
- [G]: When the maize reaches knee high (5 weeks after planting), apply top dressing fertilizer.

[51]Since the top-dressing fertilizer recommendations were not specific to the farmers' catchment area, but based on the quantity of nitrogen required to achieve a certain expected yield, specific application rates were provided to all treated farmers.

- [S]: When the maize reaches knee high(5 weeks after planting), based on soil tests around **[landmark]**, we recommend you apply 1/2 bottletop of urea per plant, making a 15 cm circle around the maize plant.
- [G/S]:Conduct second weeding 6 or 7 weeks after planting. Uproot all striga before it produces seeds because it reduces maize yields if not removed
- [G/S]: We invite you to participate in an SMS poll to help you recognize potential maize diseases and provide advice. Reply OK to start. Messages are free.
  - Do you see straight lines of holes on newly formed maize leaves?
    [if yes] This could be stalk borers. Apply insecticide e.g. bulldock or tremor, into the funnel or spay the maize plant with pentagon at top dressing.We hope this information was helpful. We will be sending another poll question tomorrow. Thank you!
    [if no] This is good news! Thank you for answering our question. We will send another question tomorrow.
  - Do you notice yellow or white streaks or discoloration on the leaves of your stunted maize plants? [if yes] It could be Maize Streak Virus. Eradicate grass weeds and use malathion or dimethoate to control as soon as possible. We hope this information was helpful. We will be sending another poll question tomorrow. Thank you!
    [if no] This is good news! Thank you for answering our question. We will send another question tomorrow.
  - Do you see striga weed in your maize plot? Striga has thin leaves and pink or purple flowers and attaches onto the maize roots.
    [if yes] Uproot all striga that has emerged. Striga competes with your maize for nutrients, water, and light and leads to reduced maize yields. We hope this information was helpful. We will be sending another poll question tomorrow. Thank you!
    [if no] This is good news! Thank you for answering our question. We will send another question tomorrow.
  - Do you see ants that cut maize stalks and feed on fallen maize cobs?
    [if yes] It could be termites. Dig out all anthills around your maize farm and ensure that you destroy the queen. Alternatively, you can dig a deep hole at the center of the anthill and use insecticide to kill the ants. We hope this information was helpful. This is the last poll question. We will NOT send another question tomorrow. Thank you for your participation!
    [if no] This is good news! This is the last poll question. We will NOT send another question tomorrow. Thank you for your participation!
- [G/S]: WEEDING REMINDER! Conduct second weeding 6 or 7 weeks after planting. Weeds compete with your maize for nutrients, water, and light, which reduces yields.
- [G]: Have you already applied top dressing fertilizer on your maize? If not, we recommend using urea at top dressing because it is a less expensive source of nitrogen.
- [S]:Have you already applied top dressing fertilizer on your maize? If not, we recommend using urea at top dressing because it is a less expensive source of nitrogen. Buy 6 kg of urea for use on 1/8 of an acre and apply 1/2 bottletop of urea per plant. Apply urea when there is enough moisture in the soil to avoid loss through evaporation.
- [G/S]: Harvest maize at physiological maturity when cobs droop and leaves dry. Dry maize in the sun even after shelling to avoid mold and attack by weevils. Maize grain must remain dry and clean during storage to avoid reduction in quantity and quality.
- [G/S]: We hope you enjoyed these messages from Precision Agriculture for Development. Our team will follow up with a phone call in the coming weeks to hear more about how your planting season went.

Additionally, during the 2017 long rains season, all treated farmers received 5 identical SMS messages about agricultural lime:

- [If pH≤5.5]: The soil in your area is [level] acidic. To avoid low yields, treat now. Apply [quantity] bottletop of lime per planting hole. [quantity] lime per 1/4 acre.
- [If pH>5.5]: The soil in your area is slightly acidic. According to our analysis, farms in your area do not need lime.

Two electronic discount coupons were sent via SMS to all participating farmers at the beginning of the season after the initial set of recommendations were sent.The first coupon was sent 10 days after the beginning of the experiment, after 7 recommendation messages, with a reminder 1 week later. The second coupon was sent 1 month after the beginning of the experiment, after 18 messages, with a reminder after 10 days and another after 20 days. All farmers, including those in the control group, received these coupons. The first coupon gave farmers a choice of either 10 kg of lime or 1 bar of soap. By allowing farmers to choose between lime and another common product of the same value, we intended to capture farmers' input choices without liquidity constraints. The second coupon, sent mid-season, provided a 30% discount on one type of top-dressing fertilizer (urea, CAN, or Mavuno), up to a pre-discount amount of 500 Ksh (approximately $ 5 USD). To redeem coupons, each farmer was assigned to an agricultural supply dealer in their preferred or closest market center (selected by farmers during baseline). To measure effects over a second season all farmers received a second round of lime coupons for the 2017 long rain season. This coupon only provided a 15% discount on the first seven 10-kg bags of agricultural lime. All farmers received a phone call around the time the coupon was sent, to ensure that treatment and controls were equally aware of the electronic coupon. A phone endline survey was conducted mid-2017 long rain season with the full sample of farmers participating in the experiment. The survey included questions about input use during the 2016 and 2017 agricultural seasons and farmers' general agricultural knowledge. Enumerators were able to complete surveys with 80% of farmers in the sample.

H.2.2 IPA/PAD2-K

The second IPA/PAD program recruited farmers through agricultural supply dealers. As part of this program, a total of 144 agricultural supply dealers in 60 market centers in Western Kenya invited farmers to enrol in a maize farmer census. The registration period ran from early December 2016 to late January 2017. A total of 8,496 farmers were registered. However, for logistical reasons the study area was later restricted to 46 market centers and 102 agricultural supply dealers. All registered farmers were then contacted over the phone by a member of the research team to obtain consent to participate in the study and baseline information about their farming practices and previous input use. A total of 5,890 farmers completed the phone baseline survey, met the eligibility criteria, and resided in eligible areas for which PAD had soil information.[52]

Farmers were then randomized into four groups. The first three groups received PAD's SMS agricultural information services and the fourth group remained as a control. One third of treated farmers received information via SMS only, another third received SMS and were invited to express interest in receiving a phone call that would explain the messages, the last third of treated farmers were contacted over the phone and offered an explanation of the messages. Messages were sent during the 2017 long rains season, and were based on ward-level soil test data.[53]

The messages focused on three types of recommendations: the use of agricultural lime

[52]From the original sample, farmers who were reached but did not complete the baseline survey included 257 who did not consent to participate in the study, 53 who were not planning to grow maize in 2017, and 40 who lived outside the four counties in which recruitment took place. Approximately 1,017 farmers lived in wards for which there was no soil test data available.

[53]The information was at the ward level. A ward is an administrative unit in Kenya. Wards were chosen because they are one of the smallest units that farmers can self-report and that soil tests could be mapped into. In western Kenya, the average size of a ward is 12 km^2.

in wards with median soil pH below 5.5, the use of planting fertilizer, and the use of top-dressing fertilizers. The soil data used to generate these recommendations was obtained by pooling data collected by 4 different organizations: IPA, OAF, Mumias Sugar Company, and the German Agro Action (Welthungerhilfe).[54] Messages were sent in either English or in Swahili, depending on farmers' language preferences at the time of registration. We list them below:

- Welcome to PAD, IPA's free advice service for maize growers. You will receive advice for your needs based on more than 10,000 soil tests from Western Kenya.
- The soil in your area is [level] acidic. To avoid low yields treat now. Apply [quantity] bottle top of lime per planting hole. [quantity] kgs for 1/4 acre. OR The soil in your area is slightly acidic. According to our soil analysis, farms in your area do not need lime.
- Soil acidity causes stunted growth.Lime reduces soil acidity and makes nutrients of DAP more available for your maize.
- When planting, apply [quantity] bottle top of lime. Cover with a handful of soil. Add [quantity] bottletop of DAP, cover with enough soil to avoid direct contact of inputs. OR When planting, apply [quantity] bottle top of DAP, cover withe enough soil to avoid direct contact of inputs.
- Check your phone! We sent you 3 planting recommendations last week [If you flash [number] before Friday this week, we will you callback soon to explain them/We will call you soon to explain them]
- Top-dress when your maize has more than 4 leaves up to knee high. If rains are good.apply 3/4 bottle top of UREA. If rains are low, apply 3/4 bottle top of CAN.
- UREA can increase your maize yields as much as CAN if rains are good. Try 11 kg of urea in 1/4 acre and see the results
- Check your phone! We sent you 2 topdressing messages this week [If you reply YES or flash [phone] by Tuesday, we will call you back soon to explain them/We will call you soon to explain them.]

A random subset of farmers also received a phone call or an SMS offer to receive a call after each set of messages explaining the content of the text messages. This 15-minute phone call did not provide any additional information, but it allowed farmers to ask clarification questions to a PAD field officer and to hear the explanation multiple times. The purpose of the phone call was to strengthen the information provided via SMS.

All farmers participating in the experiment received two electronic coupons via SMS. Each coupon allowed farmers to obtain discounts on agricultural inputs from a local agricultural supply dealer. The first electronic coupon was redeemable for 15% on the first seven 10-kg bags of agricultural lime, and the second coupon provided a 15% discount on the first 1,000 Ksh (approximately $10 USD) spent on top-dressing fertilizers (urea, CAN, or Mavuno). To avoid priming farmers about specific inputs, they were just told that the coupon would provide them with a discount for a range of agricultural inputs.[55] In addition, around 84%

[54]The IPA dataset was assembled in 2011 and 2014 in Busia county for previous projects (Fabregas et al., 2017a) and extended in 2016 as part of test plot activities in the same area. The OAF data was collected in 2016 across the entire study area. Mumias Sugar Company shared the data they collected for their operations in Busia and Kakamega counties between 2009 and 2016. The German Agro Action data was collected in Kakamega and Siaya counties in 2015.Data collected before 2014 was dropped if at least 30 more recent observations in the ward were available.

[55]To ensure that all farmers in treatment and control group were equally aware of the coupon, all farmers received a phone call a week before the program started, in which an enumerator explained how to use the coupon and at which agricultural supply dealers the coupons could be redeemed. 93% of farmers were reached during this activity.

of farmers completed a phone-based endline survey with questions about their agricultural knowledge and input use during the season.

H.3 One Acre Fund's Programs

OAF is s across six countries in Eastern and Southern Africa. In 2017, they reported working with over 600,000 farmers (OAF, 2017). The standard bundle that OAF offers includes hybrid seeds and chemical fertilizers. However, to address the problem of high soil acidity, OAF started offering farmers agricultural lime as an optional add-on. Yet, across their many locations, demand for lime remained very low. Hypothesizing that this could reflect a lack of awareness, OAF designed and evaluated several informational programs to increase lime take-up. Since OAF field officers already follow detailed protocols, a key objective of the approach was to test cheap programs that would not require additional field officer training and delivery. We describe OAF's different strategies below.

H.3.1 OAF1-K

Prior to 2016, less than 3% of OAF clients in western Kenya purchased agricultural lime through the organization (OAF, 2015). To increase take-up, OAF designed a phone-based extension pilot that consisted of six text messages targeting clients who had signed up for the OAF package during the previous season in a selected district of western Kenya.

In September 2016, during the period in which OAF farmers were placing their orders for the 2017 long rains season, OAF sent SMS messages about soil acidity and agricultural lime. Two types of messages were sent: the first, denoted as "Broad", simply encouraged farmers to use lime to reduce soil acidity and increase yields, while the second, denoted as "Detailed" provided recommendations on lime application rates and expected yield increase customized to the farmers' site. Customized messages were based on soil tests that had been previously conducted by OAF in the region.[56] In total, 4,884 farmers participated, with 3,325 farmers randomly assigned to receive messages, and 1,559 farmers remaining as a control. The same SMS message was sent six times between August and September 2016, before the OAF input contract signing period, when farmers had to decide whether to request inputs from OAF for

[56]The percentage increase in yields depended on the local level of pH and the return estimated for that pH level based on OAF farm trials.

the following season.

Farmers in both treatment groups received 6 identical messages, all messages were sent in Swahili. We report the messages below:

- [Broad]: Hello [name],Your soil is acidic. Use lime to reduce acidity and increase yields.Call xxx-xxxx.
- [Detailed]: Hello [name],Your soil is [level] acidic. We recommend [amount] kg of LIME per acre at [total cost] Ksh. Use lime to reduce acidity and increase yields [percentage increase]%.Call xxx-xxx.

To measure outcomes we use two sources of data OAF administrative data and phone survey data collected by researchers. The administrative data contains information on loan enrollment and inputs purchased through the OAF program for the 2017 and 2018 long rain seasons. However, only 60% of farmers who received the text messages went to receive OAF loans in the 2017 long rain agricultural season. This outcome is an imperfect measure of the overall effects of the program on lime purchases if farmers acquired lime from other sources. To explore this possibility and obtain additional information from farmers, a follow-up phone survey led by IPA was conducted in May 2017 with a random sample of 30% of the farmers participating in the trial. This survey asked respondents about their knowledge of lime and their input use during the 2017 long rains season. About 79% of selected farmers were surveyed.

In September 2017, at the time of enrollment for the 2018 long rains season, a subset of the farmers who purchased inputs from OAF for the 2017 long rains season received additional messages encouraging lime adoption.The treatment assignment for this program was stratified on previous treatment status.

H.3.2 OAF2-K

A second OAF program was implemented with approximately 30,000 farmers in four Kenyan districts in September 2017. Former OAF clients were randomized into a no message control group or a treatment group receiving SMS messages encouraging lime adoption (which did not depend on results from soil tests in the area). Additionally, a quarter of farmers were randomly assigned to receive additional messages encouraging the use of additional fertilizer (Extra CAN) for top-dressing.

Six different types of messages were sent: a “Basic” message simply recommended to purchase lime, a message, “Yield increase”, also mentioned that lime would increase yields, two encouraged experimentation, "Experimentation (selfish)" and "Experimentation (neighbors)", and two leveraged on behavioral nudges "Social comparison" and "Self-efficacy". Half of the treated farmers were randomly assigned to receive messages addressing the whole family instead of the individual (by replacing the word "you" with "your family"). The messages encouraging use of additional quantities were identical to those encouraging use of lime (the word "Lime" was replaced by "Extra CAN"). Farmers assigned to receive both lime and fertilizer message were randomly assigned to receive one of the two first and the other on the next day for all repetitions. The number of repetitions (from 1 to 5) and the frequency of the messages (every 2, 4, 6, or 8 days) were cross-randomized. In September 2018, at the time of enrollment for the 2019 long rains season, all the farmers who purchased inputs from OAF for the 2018 long rains season received additional messages encouraging lime adoption (but no messages about fertilizer). We report all messages below:

- [M1: Basic] [Name], OAF recommends [you/your family] register to buy [Lime/Extra CAN] for your maize.
- [M2: Yield increase] [Name], OAF recommends [you/your family] register to buy [Lime/Extra CAN] for your maize. You'll get higher yields by using [Lime/Extra CAN].
- [M3: Experimentation (selfish)] [Name], OAF recommends [you/your family] register to buy [Lime/Extra CAN] for your maize. Try it on just a small part of your land to see the benefits.
- [M4: Experimentation (neighbors)] [Name], OAF recommends [you/your family] register to buy [Lime/Extra CAN] for your maize. Try it on just a small part of your land to so that you and your neighbors can see the benefits.
- [M5: Social Comparison] [Name], OAF recommends [you/your family] register to buy [Lime/Extra CAN] for your maize. Farmers all over Western are getting bigger yields by using [Lime/Extra CAN]. Keep up with them!
- [M6: Self-efficacy] [Name], OAF recommends [you/your family] register to buy [Lime/Extra CAN] for your maize. You have the ability to achieve higher yields by using [Lime/Extra CAN]!

Farmers were later matched to OAF administrative data to measure their likelihood of demanding agricultural lime and other inputs for the following two agricultural seasons. About 76% of farmers who received text messages decided to acquire any inputs through OAF that season. Again, we define the primary outcome variable as the probability of purchasing agricultural lime from OAF.

H.3.3 OAF3-R

OAF-Rwanda, known as Tubura, implemented an SMS-based program aimed at encouraging experimentation with a type of agricultural lime, known in Rwanda as travertine.[57] The messages were sent in June 2017, when farmers were enrolling for the 2018 main agricultural season (September 2017 to January 2018). In Rwanda, OAF partners with the government to provide goods, services, and training to rural farmers. Since 2016, OAF and the government of

[57] To simplify the exposition we consistently use the terms OAF and agricultural lime.

Rwanda have engaged in a concerted effort to promote adoption of agricultural lime. Unlike the Kenya OAF program that operates only during the main agricultural season, the OAF Rwanda provides inputs on credit for both seasons. Farmers who want to purchase inputs for the secondary season (February to August) need to place their orders before the beginning of the previous main agricultural season (September to January), but are allowed to drop some products before the time of delivery.

As for OAF2-K, the purpose of this program was to understand how to optimize message content, framing, number of repetitions and framing. Given the relatively low mobile phone penetration in the country, OAF wanted to explore ways to increase spillovers within farmers' group in order to reach all farmers. For this reason, the first stage of randomization took place at the group level, assigning farmers groups to four group-level treatments: a pure control group where no farmers in the group, a treatment in which all farmers received messages, and a treatment in which farmers received messages with probability 0.5 and content and framing were randomized at the individual level. From a total of 216,475 farmers registered in the OAF program, 114,569 had a phone registered in the database, and 85,160 had a unique phone number. In our analysis we drop all farmers that did not have a phone registered in the database and consider the original treatment assignment, regardless of whether phones are shared or not. The main results are robust to excluding all farmers with shared phones from the analysis. Messages are listed below:

- [T1-G: General promotion (gain)] Many fields in Rwanda have acidic soil and need TRAVERTINE to increase yields. Order from TUBURA now.
- [T1-L: General promotion (loss)] Many fields in Rwanda have acidic soil and need TRAVERTINE to avoid a yield loss. Order from TUBURA now.
- [T2-G: Specific+ yield impact (gain)] Many fields in Rwanda have acidic soil. Applying 25 kg/are of TRAVERTINE will increase yields by 20%.Order from TUBURA now.
- [T2-L: Specific+ yield impact (gain)] Many fields in Rwanda have acidic soil. Applying 25 kg/are of TRAVERTINE will prevent a yield loss of 20%. Order from TUBURA now.
- [T3-G: Self-diagnosis (gain)] Do you have fields with poor harvests even when you use fertilizer? You probably have acidity and need TRAVERTINE to increase yields. Order from TUBURA now.
- [T3-L: Self-diagnosis (loss)] Do you have fields with poor harvests even when you use fertilizer? You probably have acidity and need TRAVERTINE to avoid a yield loss. Order from TUBURA now.
- [T4-G: Soil test (gain)] Ask your Field Officer for a free soil test to learn if your fields are acidic and you need to order TRAVERTINE to increase yields.
- [T4-L: Soil test (loss)] Ask your Field Officer for a free soil test to learn if your fields are acidic and you need to order TRAVERTINE to avoid a yield loss.
- [T5-G: How travertine works (gain)] Many fields in Rwanda have acidity, which blocks fertilizer uptake. Applying TRAVERTINE solves the problem, increasing crop yields. Order from TUBURA now.
- [T5-L: How travertine works (loss)] Many fields in Rwanda have acidity, which blocks fertilizer uptake. Applying TRAVERTINE solves the problem, preventing a yield loss. Order from TUBURA now.
- [T6-G: Order immediately (gain)] Many fields in Rwanda have acidic soil and need TRAVERTINE to increase yields. Order it immediately, when signing your TUBURA order form.
- [T6-L: Order immediately (loss)] Many fields in Rwanda have acidic soil and need TRAVERTINE to avoid a yield loss. Order it immediately, when signing your TUBURA order form.
- [T7-G: Your cell is acidic + yield impact (gain)] In your cell the soil is acidic. If you apply 25 kg/are of TRAVERTINE you can boost yields by 20%. Order from TUBURA now.
- [T7-L: Your cell is acidic + yield impact (loss)] In your cell the soil is acidic. If you apply 25 kg/are of TRAVERTINE you can avoid a yield loss of 20]%. Order from TUBURA now.

Social nudge message:

- [SN] Please share this information about TRAVERTINE with your group members and neighbors, especially those who don't have phones!

We can measure whether farmers purchased lime from OAF for the 2018 agricultural seasons (about 65% of control farmers enrolled in the OAF loan program for the main agricultural season).

I Local Agricultural Recommendations

In this section we briefly describe how the local agronomic recommendations were constructed.

IPA-K/PAD1-K. Lime recommendation for IPA-K farmers where calculated based on median level of exchangeable acidity in the area. Since exchangeable acidity information was not available for the MSC sample, lime recommendations were based on the median level of pH in the farmers' fields.

The amount of planting fertilizer recommended was based on the median amount of phosphorus measured in the area which determined the recommended quantity of diammonium phosphate (DAP).

Top-dressing fertilizer recommendations were based on the quantity of nitrogen required to achieve a certain expected yield. The quantity was selected based on the target yield of 2 t/ha. For this target yield the quantity of nitrogen required is 54 kg per hectare, which corresponds to 117 kg of urea or 206 kg of calcium ammonium nitrate (CAN). Given the lower amount required and the fact that the average price of urea was lower than that of CAN (66 Ksh vs 76 Ksh per kg) urea was recommended.

IPA-K/PAD2-K. Farmers participating in this study were provided with lime and planting fertilizer (DAP) recommendations based on the local level of soil acidity, measured in terms of soil pH, and phosphorus, respectively. The recommended input quantities were based on a target yield of 2 t/ha, which represents an improvement with respect to the baseline average of 1.42 t/ha, while keeping the cost of the input package affordable for the farmers.

The amount of lime recommended was decreasing in the level of pH while the amount of DAP recommended[58] was decreasing in the level of phosphorus.

The recommendations are based on micro-dosing, rather than general broadcasting meth-

[58] The government's recommended application of phosphorus for Western Kenyan soils, for a target yield of 3.9 t/ha in soils with P below 10 mg/kg, is 26kg P/ha, corresponding to 130 kg DAP/ha, (FURP, 1995; Wasonga et al., 2008). With a target yield of 2 t/ha, the recommendations provided as part of this study involved applying 21kg P/ha, corresponding to 107 kg DAP/ha.

ods, to maximize effectiveness[59] (IPNI, 1999). To provide a standard measure for micro-dosing recommendations, farmers were advised to use a soda bottletop, which is a common item easily available throughout the study area.

Recommendations for lime and DAP were provided based on median soil characteristics in the farmers' ward.[60] Soil data was pooled from 5 different sources: (1) Soil data collected by IPA-K in Busia county for previous projects (Fabregas et al., 2017a) in 2011 and 2014 and as part of test plot activities conducted in 2016. (2) Soil data collected by One Acre Fund across the entire study area in 2016. (3) Soil data collected by Mumias Sugar Company in Busia and Kakamega counties between 2009 and 2016. (3) Soil data collected by the German Agro Action (Welthungerhilfe) in Kakamega and Siaya counties in 2015.

These datasets provided over 30,000 soil tests for the area in which the study took place, specifically, the set of wards in which the farmers participating in the intervention are based. However, in order to base the recommendations on the most recent data, data was dropped for soil tests performed before 2014, when possible[61] The final dataset used included about 7,085 observations for 108 wards.

OAF1-K. The standard application rate recommended by OAF and reflected in field materials was 200kg/acre across the entire program. In order to generate "local" recommendations OAF's used their own own soil tests, performed using soil spectroscopy, and soil data collected for a previous project by IPA-K (Fabregas et al., 2017b) and analyzed by the Kenya Agricultural Research Institute (KARI) using wet chemistry in 2011 and 2014. These soil chemistry results were then interpolated across areas through Kriging to create a continuous field of soil chemistry predictions. Since OAF does not collect the coordinates of farmers' plots, farmers were assigned to the GPS coordinates of the site to which inputs are delivered by OAF.

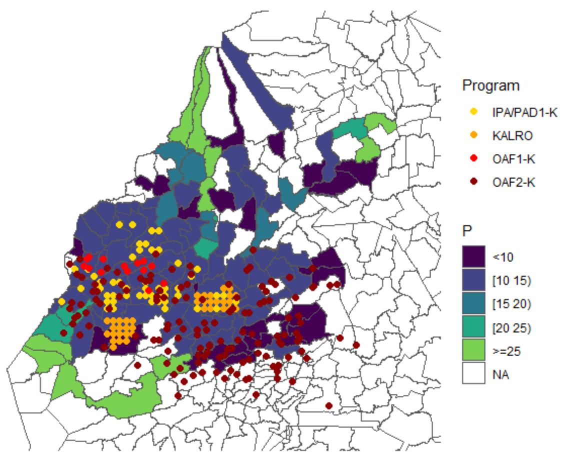
[59] Farmers were also recommended to use micro-dosing for lime application as it requires lower investment and yields higher returns in the short term (Mortvedt and Follet, 1999; Terman and Engelstad, 1976; Plaster, 2003; OAF, 2015). This practice is not recommended by the local government and the Kenya Agriculture and Livestock Research Organization, which recommend broadcasting. However, it is in line with the recommendations of One Acre Fund, an NGO that serves about 400,000 farmers in the region, including 35% of the farmers in our sample.

[60] Recommendations were provided based at the ward level because that is the most precise information collected about farmers' location. The data was aggregated into medians because the majority of the soil data available was not geocoded and only provides information on the administrative unit in which the sample was collected.

[61] Data collected before 2014 was dropped if at least 30 more recent observations in the ward were available. Since the data displays clear trends of decreasing pH and phosphorus levels over time, they were adjusted using coefficients based on the Mumias Sugar Company soil data: a coefficient of -0.027 per year was applied for pH and -0.504 per year for phosphorus. These coefficients were obtained by regressing pH and phosphorus data on a time trend and constant, controlling for field fixed effects, these regressions are based on a sample of over 60,000 observations.

Optimal lime application rates, for each level of pH, were based on OAF on-farm agronomic trials conducted in 2015 (OAF, 2015). During that trial three different lime application rates were tested: 50kg/acre, 100kg/acre and 200kg/acre. The sample was divided according to pH quintiles and, for each quintile, the lime application rates that resulted in the most precisely estimated effect on yield was chosen. Two different lime application rates were recommended, based on the local predicted level of pH: 200kg/acre and 50kg/acre.

Figure I1: Soil Map of Western Kenya



(a) Phosphorus

Notes: Panel (a) shows the median level of Phosphorus (P) in all wards in which the IPA/PAD2-K program took place as well and the location of the other programs.