American Economic Journal: Applied Economics 2025, 17(1): 527–566
https://doi.org/10.1257/app.20220072

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

By Raissa Fabregas, Michael Kremer, Matthew Lowes, Robert On, and Giulia Zane*

While some studies suggest mobile phone–based information programs change behavior; others find no effect. We evaluate six text message agricultural extension programs, collectively covering 128,000 farmers. A meta-analysis finds a 1.22-fold increase in the odds of adoption of recommended practices (95 percent CI: 1.16, 1.29). We cannot reject similar impacts across experiments. Impacts are increased by message repetition, but not by providing more granular information, using behavioral framings, or complementing texts with phone calls. There is little evidence of message fatigue or crowd-out. Despite modest absolute impacts detectable only with large samples or meta-analysis, texts are inexpensive enough to be cost-effective. (*JEL* D83, D91, L96, O13, Q12, Q16)

The widespread adoption of mobile phones in developing countries over the past few decades has opened up new avenues 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 barriers and change individual behavior (GSMA 2020). While only a fraction of these initiatives have been evaluated, there is a growing literature assessing the

* Fabregas: University of Texas at Austin (email: rfabregas@utexas.edu); Kremer: University of Chicago, and NBER (email: kremermr@uchicago.edu); Lowes: Independent Researcher (email: matt.lowes@gmail.com); On: The Agency Fund (email: robert@agency.fund); Zane: International Water Management Institute (email: g.zane@cgiar.org). Seema Jayachandran was coeditor for this article.We thank our partner organizations: the Kenya Agriculture and Livestock Research Organization, and in particular Dr. Martins Odendo, One Acre Fund, Precision Development, and Innovations for Poverty Action. We are also grateful to the Mumias Sugar Company, and German Agro Action (Welthungerhilfe) for sharing soil information and specially to Dr. Javier Castellanos and to Dr. David Guerena. Lillian Alexander, Carolina Corral, Tomoko Harigaya, Charles Misiati, Chrispinus Musungu, Cara Myers, Alex Nawar, Carol Nekesa, William Wanjala, Violet Omenyo, Victor Perez and Megan Sheahan provided excellent project support. This project also benefited from feedback from Rema Hanna, Rohini Pande, Kashi Kafle, Jessica Zhu, Jack Marshall, Andrew Dustan, Luiza Cardoso, Rachel Meager and Witold Wiecek and seminar or conference participants at AFE, MWIDC, the Texas Experimental Association Symposium, NEUDC, Sao Paulo School of Economics, Development Innovation Lab, Y-RISE, McGill, and JILAAE. Finally, we are grateful to the journal editor and the anonymous referees for many suggestions that improved the paper. Funding from 3ie (TW4.1101), ATAI (5710004168), the Dioraphte Foundation, Unorthodox Philanthropy, PxD, and an anonymous donor made this project possible. Disclosure: Kremer is a board member of PxD. Zane's postdoctoral fellowship was funded by PxD. Fabregas, Kremer, and On did not receive any financial compensation from PxD, but the organization partly funded this research. On volunteered for 1AF at the time of the experiment. Lowes worked for 1AF. A high-level summary of the results presented in this paper appears in Fabregas, Kremer, and Schilbach (2019), which cites a working version of the article. The project was approved by the Harvard IRB (IDs: IRB16-1121, IRB13-3434, IRB17-06991), and is registered at the AEA RCT Registry (AEARCTR-0006837) (Fabregas et al. 2020a).
† Go to https://doi.org/10.1257/app.20220072 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

effectiveness of these programs across a range of sectors, from health (Hall et al. 2014; Jamison, Karlan, and Raffler 2013), education (Aker, Ksoll, and Lybbert 2012; Cunha et al. 2017; Angrist, Bergman, and Matsheng 2022) and finance (Karlan, Morten, and Zinman 2015; Karlan et al. 2016) to governance (Dustan, Hernandez-Agramonte, and Maldonado 2023; Buntaine et al. 2018; Grossman, Humphreys, and Sacramone-Lutz 2020) and agriculture (Aker, Ghosh, and Burrell 2016; Fafchamps and Minten 2012; Cole and Fernando 2021).

The empirical evidence on the impacts of these programs on recipient behavior has been characterized as mixed (Aker, Ghosh, and Burrell 2016; Deichmann, Goyal, and Mishra 2016; Baumüller 2018; Grossman, Humphreys, and Sacramone-Lutz 2020; Steinhardt et al. 2019). If program effectiveness is very sensitive to the specific features of its design, the identity of the implementing organization, targeted recipients, or the local context, it might be difficult to draw broader policy conclusions about whether to scale up or extend these interventions to a new setting (Pritchett and Sandefur 2015). However, the perception of mixed results could also stem from sampling variation (Meager 2019), as well as selection biases (Glewwe et al. 2004), varying levels of statistical power (Ioannidis, Stanley, and Doucouliagos 2017), differences in measurement, or publication biases (DellaVigna and Linos 2022).

This paper examines the role of digital interventions on behavior change by presenting new experimental evidence on the impacts of six text message–based agricultural extension programs on individuals' decisions 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 adopted. Despite this potential, texting might be too impersonal, light-touch, or restrictive to meaningfully convey information. Illiteracy, mistrust, or mistargeting might also limit the effectiveness of these types of programs (Aker 2017), especially when implemented at scale (Bird et al. 2019).

The programs examined in this study were implemented in Kenya and Rwanda by three different organizations: a public agency, a social enterprise, and a research-oriented nonprofit. All the programs shared the goal of increasing farmer experimentation with locally recommended agricultural inputs. Despite sharing similar objectives and using mobile phones to reach out to farmers, 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 setup allows us to estimate impacts for each program individually and aggregate the results through a meta-analysis. The meta-analysis increases statistical power and enables formal testing of impact heterogeneity across studies. This configuration also captures a common occurrence in program implementation: organizations with similar tools and objectives often design and adapt their programs differently based on their specific constraints, philosophies, and opportunities. When considering scalability, it is important to understand to what extent these implementation 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 128,000 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, Stanley, and Doucouliagos 2017; Dahal and Fiala 2020; McKenzie and Woodruff 2014). This interpretation is particularly problematic for very cheap interventions, such as text messages, since the effect sizes required for these programs to be cost-effective are usually very small. [1]

Second, across all experiments, we use actual input acquisitions as our primary measure of behavior change. To track the input purchases of farmers affiliated with the social enterprise, we use the organization's administrative data, as they were directly selling inputs to farmers. For projects targeting independent farmers, we rely on the redemption of input discount coupons by both treatment and control farmers in dozens of small agricultural shops in the region. Using data from real purchasing decisions mitigates the risk that any estimated effects are driven by social desirability or courtesy bias in self-reports, a common concern in the evaluation of informational programs (Baumüller 2018; Haaland, Roth, and Wohlfart 2023). Using survey endline data for four programs, we can also compare self-reports against the administrative records and investigate other outcomes such as increased knowledge, adoption of other recommended practices, as well as any potential crowd-out in the use of nonrecommended inputs.

Combining the effects of all six programs in a meta-analysis using odds ratios (OR), a relative measure of effects that is less sensitive to variations in baseline input use probabilities, we find a small but statistically significant effect on following the recommendations (OR: 1.22, 95 percent CI 1.16 to 1.29, $N = 6$). The aggregate effect for following recommendations about a newly introduced technology (agricultural lime) is 1.19 (95 percent CI 1.11 to 1.27, $N = 6$), whereas the effect for following recommendations for largely unused types of a well-known technology (chemical fertilizers) is 1.27 (95 percent CI: 1.15 to 1.40, $N = 4$). With only six experiments, we cannot draw definitive conclusions about the extent of program heterogeneity. We do observe that some individual experiments had statistically significant impacts and others did not. However, we cannot reject the hypothesis that the effects were the same and that the observed differences may primarily stem from sampling variation.

Reestimating the meta-analysis using an absolute effect measure derived from linear probability models yields a statistically significant increase of 2 percentage points (95 percent CI 0.01 to 0.03, $N = 6$) on the probability of following the recommendations. Using this effect measure we reject the null hypothesis of homogenous treatment effects across programs. However, the variation in point estimates across different experiments is small, typically within a range of 1 or 2 percentage points. Together with analysis of heterogeneity in treatment effects using observed correlates in micro data, this observation aligns with the notion that true impact heterogeneity across experiments is limited, at least, beyond the variation that can be attributed to initial levels of input adoption.

[1] In 2020, some services in Kenya charged less than \$0.003 per text message. In India, it varied from \$0.0004 to \$0.006, 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.

Although the programs were primarily intended to deliver new information to farmers rather than just act as reminders or nudges, the effects seem to operate partly through behavioral channels rather than purely through learning. On the one hand, treated farmers were significantly more likely to correctly identify the purpose of the newly introduced input (OR: 1.53, 95 percent CI 1.38 to 1.70, $N = 4$). On the other hand, the effects on input purchases waned after one season, but re-treatment helped sustain impacts. Moreover, we cannot reject that the impacts on input use were the same regardless of farmers' baseline levels of knowledge about these technologies. A potential interpretation of these results is that well-timed messages might be effective, partly because they affect how top-of-mind a decision is (Bettinger et al. 2021; Karlan, Morten, and Zinman 2015; Karlan et al. 2016; Raifman et al. 2014).

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, Woodman, and Chatterji 2015; Karlan and Zinman 2008), we find larger effects on the use of the newly introduced input when estimated using survey data compared to administrative purchase data. We cannot definitively attribute these differences to misreporting, since farmers might have acquired inputs from different sources. However, within the context of one project for which we have more information, farmers for whom there was a mismatch in the survey and administrative data were less able to report the shop they had acquired the input from and more likely to report acquiring the input from sellers who had not stocked it during the same period. This highlights the risk of relying on self-reports to assess behavior change, even when the reported behavior is not particularly sensitive, and points to the importance of collecting objective data. Further, it is difficult to predict the direction of these inconsistencies; the discrepancy affects some but not all experiments, and we do not find significant differences in estimated impacts for fertilizers.

Finally, we use individual project experimental variation to draw additional lessons about the importance of different programmatic features. We cannot reject the hypothesis that messages crafted using behavioral insights (e.g., sense of urgency, self-efficacy, social comparisons, etc.) were no more effective than basic messages. 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. The lack of additional impacts resulting from providing more granular information aligns with experimental findings from other contexts (Beg, Islam, and Rahman 2024; Corral et al. 2020).

To test whether person-to-person communication could strengthen the effects, one experiment complemented the text messages by randomizing a phone call from a field officer. We find no additional effects from this add-on. Message repetition, however, was modestly effective at increasing purchases. We also estimate spillovers for the programs that targeted users who belonged to farmer groups, and find some suggestive evidence of positive externalities.

Without data on yields, we cannot draw firm conclusions on the effects on farmers' profits.[2] However, using agronomic estimates of the impacts of the recommended

[2] Measuring impacts on downstream outcomes, such as yields and profits, can be complex especially if the effect sizes that would make these programs cost-effective are very small. Self-reported yields are noisy (Lobell

inputs on maize yields, we provide back-of-the-envelope calculations of the potential benefits of sending these messages. Our estimates suggest that the benefit-cost ratio for a similar program is about 46:1 if operated at a sufficiently large scale such that the per-farmer fixed costs become negligible. Text messages can also compare favorably to more intensive, but also more expensive, programs such as in-person farmer events.

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 2022). In our case, the focus is on digital informational interventions in low-income contexts. Our results also speak to the literature concerned with the use of experimental evidence for policy scale-up, where heterogeneity in treatment effects has been used as a measure of external validity (Pritchett and Sandefur 2015; Allcott 2015; Meager 2019; Vivalt 2020). We find little evidence to support the notion that true program impacts are highly heterogeneous, and we suggest caution in qualitatively interpreting differences in statistical significance across studies because these differences could be driven by sampling variation and studies underpowered to detect small effects.

Finally, we complement expert qualitative summaries on digital interventions for development (Aker 2017; Aker, Ghosh, and Burrell 2016) and the handful of experimental studies assessing the impacts of digital agricultural extension systems. A text-based program for Ecuadorian farmers increased knowledge and self-reported adoption of integrated soil management practices (Larochelle et al. 2019). Text messages sent by an agribusiness to sugar cane farmers in Kenya had positive yield impacts in one trial but no statistically significant effects in a second trial (Casaburi et al. 2014). Fafchamps and Minten (2012) report null effects from a text-based program with weather, price, and advisory content in India. A voice-based service, targeted at cotton farmers in India, increased the reported use of recommended seeds (Cole and Fernando 2021). This paper expands what is known empirically about text-based agricultural extension programs, addressing some methodological limitations in existing work.[3]

This paper is organized as follows: Section I presents the context and design of each program and their evaluations. Section II discusses the empirical strategy. Section III presents the main results, and Section IV discusses some of the additional lessons that we can be drawn from individual experiments. We present cost-effectiveness estimates in Section V and conclude in Section VI.

et al. 2020), and objective measures such as physically harvesting a section of a farmer's plot can 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).

[3] A few additional literatures are worth mentioning. First, research on the broader effects of mobile phone access on market performance and productivity (Jensen 2007; Gupta, Ponticelli, and Tesei 2020; Aker and Mbiti 2010; Aker and Fafchamps 2015). Second, studies on the effects of providing crop prices through phones (Camacho and Conover 2019; Mitra et al. 2018; Nakasone, Torero, and Minten 2014; Courtois and Subervie 2015; Svensson and Yanagizawa 2009). Finally, delivering information via video, tablets, or smartphone apps, have also been shown to have positive effects on farmers' beliefs and behaviors (Tjernström et al. 2021; Van Campenhout, Spielman, and Lecoutere 2021; Arouna et al. 2021). However, until smartphone penetration increases, such approaches might be difficult to scale.

I. Context, Programs, and Experimental Design

A. *Context*

The programs targeted maize smallholder farmers in Rwanda and Kenya between 2015 and 2019 (see online Appendix Figure A1 for a map).[4] In both countries, maize is farmed twice a year. In Kenya, the long rains season (the primary agricultural season), runs from March to August, and the short rains season (the secondary season), from September to December. The main season in Rwanda is from September to January, and the secondary season from March to August.

In these areas, 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 can dramatically reduce crop yields by limiting nutrient availability to plants (The et al. 2006; Tisdale, Nelson, and Beaton 1990; Brady and Weil 2004). Soil pH is considered optimal for crop growth at around 6.0–7.0. Soil pH is considered strongly acidic when under 5.5, and this pH level is a standard threshold under which the soil is deemed unsuitable for maize growth (Cranados, Pandey, and Ceballos 1993; Kanyanjua et al. 2002). The application of agricultural lime is one of the cheapest and most widely recommended methods to increase pH. Multiple public agencies and NGOs in Africa recommend the use of lime on acidic soils, and experimental plots conducted in Kenya suggest that lime application can increase maize yields by 5–75 percent depending on the area, soil characteristics and rate applied (Kisinyo et al. 2015; Gudu et al. 2005; 1AF 2014). Yet agricultural lime is not a widely known or used input. In Kenya, only 7 percent to 12 percent of farmers in our samples reported having ever used it at baseline, and in Rwanda, only 6 percent had purchased it during the previous season.

Chemical fertilizers are more widely used, but most farmers in the sample areas have only used a specific phosphate-based fertilizer, diammonium phosphate (DAP), applied at planting time. Few farmers regularly experiment with other options, such as top-dressing fertilizers like calcium ammonium nitrate (CAN) and urea, which are applied to the plant once it has started to mature.[5] Fertilizers, particularly top-dressing ones, can be profitable (Duflo, Kremer, and Robinson 2008; Kelly and Murekezi 2000) and current national and international recommendations have started to encourage farmers to use fertilizers that best fit their soil and local conditions (KSHC 2014; NAAIAP 2014). Therefore, several organizations aim to inform farmers about different fertilizer options to encourage experimentation.

[4] The adult literacy rate in Kenya is 82 percent and in Rwanda is 73 percent (UNESCO 2022). Estimates of mobile phone penetration in 2017 were 87 percent for Kenya and 48 percent in Rwanda, though significant urban-rural gaps exist (Gillwald and Mothobi 2019).

[5] In Kenya, over 80 percent of farmers in our sample reported using planting fertilizers in the previous seasons. This proportion is higher than that reported in other contexts in the region (Sheahan and Barrett 2017). Indeed, input use in Kenya is also much higher than in neighboring countries. In 2019, the FAO estimated that Kenya's average nitrogen fertilizer use was 22 kg, which is similar to that of Ethiopia (23 kg) but higher than that of Uganda (1.2 kg) or Rwanda (7 kg).

One potential reason for the low uptake and experimentation with locally suitable technologies is that many smallholders lack reliable access to science-based agricultural advice. Access to extension services is rare for farmers not engaged in agricultural programs promoted by NGOs or other organizations. Acquiring information about locally relevant inputs is not trivial, even if farmers find this advice valuable (Fabregas et al. 2020b). For instance, there might be significant frictions in information sharing among peers (Chandrasekhar et al. 2022), and learning through self-directed experimentation can be difficult if farmers do not know over what dimensions to experiment (Hanna, Mullainathan, and Schwartzstein 2014) or if they misperceive their soil characteristics (Berazneva et al. 2016). Moreover, potential heterogeneity in soil characteristics and the profitability of inputs (Marenya and Barrett 2009; Suri 2011) makes it difficult to rely on national or regional-level blanket recommendations.

Knowledge gaps are apparent in the data. For example, in our initial evaluation in Kenya, at baseline, 22 percent of farmers were unsure if soil acidity was an issue, while 40 percent considered it a significant problem for their soil. Among those who considered it a significant problem, 67 percent were unaware of any methods to address it, and only 4 percent were familiar with agricultural lime as a potential solution. Combining all samples for which we have data on farmers' knowledge, we estimate that only 32 percent of untreated farmers were able to recognize lime as a potential solution to high soil acidity.

B. *Partner Organizations*, *Programs*, *and Randomization*

This section summarizes the characteristics of the implementing organizations, their programs, and features of each evaluation (Fabregas et al. 2020a).[6] 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 programs, and Table 2 describes the characteristics of each experiment. Details about each program and evaluation are discussed in online Appendix J.

KALRO.—The Kenya Agriculture and Livestock Research Organization (KALRO) is a semi-autonomous public agency with the mandate to promote agricultural research and dissemination. KALRO's text message program was developed in partnership with the Ministry of Agriculture and was envisioned as a low-cost way to reach farmers. In total 21 messages were sent throughout the 2015 short rains season, including two messages pertaining to lime and three messages focused on fertilizers.

Participating farmers were recruited by field agents who went door-to-door in KALRO's catchment areas. Among visited farmers, 95 percent met the inclusion criteria for the study (i.e., phone owner, responsible for farming, growing maize during the previous season) and were invited to complete a baseline survey. Farmers were then randomized at the individual level into a treatment or a control arm.[7] All invited

[6] We define an "implementing organization" as the primary organization designing the programs and crafting and delivering the messages. Each implementing organization faced its own constraints, goals, and directives. IPA and PxD-affiliated researchers were involved in evaluating all six programs.

[7] A second treatment arm, testing in-person farmer field days, was also evaluated as part of this project. The results are described in Fabregas, Kremer, and Schilbach (2023).

Table 1—Program Characteristics

	KALRO (1)	IPA/PxD1-K (2)	IPA/PxD2-K (3)
Org.	KALRO	IPA/PxD	IPA/PxD
Org. type	Public	NGOs	NGOs
Location	Kakamega and Siaya (Kenya)	Busia and Kakamega (Kenya)	Busia, Bungoma, Kakamega, and Siaya (Kenya)
Agricultural season	SR 2015	SR 2016/LR 2017	LR 2017
Recruitment	Farmers drawn from village census survey	Former NGO and contract farming participants	Clients of agrodealers
Eligibility	Farmed during past year, in charge of farming	Planted maize in prior season, reside in program area	Clients of agrodealers
Message content	Lime, fertilizer, seeds, field management	Lime, fertilizer, field management	Lime and fertilizer
Number of messages	21 total (2 acidity/lime; 3 fertilizer)	24–28 total (7–9 acidity/lime; 4–9 fertilizer)	13 total (6 acidity/lime; 4 fertilizer)
Timing	Throughout season	Throughout season	Before planting and topdressing
Lime recommended?	All (if acidic)	0.82	0.77
Key fertilizers recommended	DAP, NPK, CAN, Mavuno	Urea	Urea
Used local soil data?	No	Yes	Yes
Additional services?	No	No	Phone-call
Any message repetition	No	Yes	Yes
Opt-in	1	0.95	0.95
Previous lime use[b]	0.07	0.12	0.09
Previous season fert. use (any/recommended)[b]	0.84/0.84	0.93/0.18	0.88/0.19
Female[b]	0.65	0.37	0.34
Primary school[b]	0.53	0.63	0.71

	1AF1-K (4)	1AF2-K (5)	1AF3-R (6)
Org.	1AF	1AF	1AF
Org. Type	Social Enterprise	Social Enterprise	Social Enterprise
Location	Busia and Kakamega (Kenya)	Bungoma, Busia, Kakamega, and Vihiga (Kenya)	Western, Eastern, Southern (Rwanda)
Agricultural season	LR 2017/LR 2018	LR 2018	Main season 2018/2019
Recruitment	1AF clients in LR 2016	1AF clients in LR 2017	1AF clients in 2017
Eligibility	1AF clients in LR 2016	1AF clients in LR 2017	1AF clients in 2017
Message content	Lime	Lime and fertilizer	Lime
Number of messages	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)
Timing	Before input choice	Before input choices	Before input choice
Lime recommended?	All	All	All
Key fertilizers recommended	–	CAN	–
Used local soil data?	Yes	Yes	Yes
Additional services?	1AF Services and call-center	1AF Services and call-center	1AF Services and call-center
Any message repetition	Yes	Yes	Yes
Opt-in	–	–	–
Previous lime use[b]	–	–	0.06
Previous season fert. use (any/recommended)[b]	0.95/–	0.93/0.15	0.95/–
Female[b]	0.65	0.69	–
Primary school[b]	–	–	–

Notes: SR denotes Short Rains Season and LR Long Rains Season. [b] denotes data for the 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. "Key fertilizer recommended" denotes whether fertilizer messages were sent, and if yes, the types of fertilizer recommended. "Opt-in" indicates the fraction of farmers who, when invited, agreed to receive texts. "Previous season fert. use (any/recommended)", for KALRO and IPA/PxD refers to whether farmers report using any chemical fertilizer (any) and the fraction that used they key fertilizer programs recommended during the previous LR season. For 1AF "any" refers to farmers that reported to 1AF planting maize (the standard package includes fertilizers) and "recommended" refers to the fraction that purchased extra CAN from 1AF during previous LR season. Baseline data refers to the following LR or main seasons: 2014 (KALRO), 2015 (IPA/PXD1-K), 2016 (IPA/PXD2-K), 2016 (1AF1-K), 2017 (1AF2-K), 2017 (1AF3-K). A requirement across all programs was to have access to a mobile phone.

Table 2—Research Design

	KALRO (1)	IPA/PxD1-K (2)	IPA/PxD2-K (3)
Unit of randomization	Individual	Individual	Individual
Sample size	773	1,897	5,890
Treatment arms (#)	1	2	3
Treatment arms	**1. SMS**	**1. General SMS**, **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**: also offered to receive phone call.
Second season text	No	Yes, maintain treatment status	No
Admin. outcome	Coupon (paper), LR 2016	Coupon (digital), SR 2016 and LR 2017	Coupon (digital), LR 2017
Coupon value	50 percent discount lime, 50 percent discount td fertilizer	Choice 10 Kg lime or soap (first season); 15 percent discount lime (second season); 30 percent discount CAN or urea	15 percent discount lime; 15 percent discount fertilizer
Baseline survey	Yes	Yes (phone)	Yes (phone)
Endline survey	Yes, SR 2015	Yes (phone), LR 2017	Yes (phone), LR 2017/SR 2017

	1AF1-K (4)	1AF2-K (5)	1AF3-R (6)
Unit of randomization	Individual	Individual	Cluster (farmer group)
Sample size	4,884	32,572	82,873 (17,850 groups)
Treatment arms (#)	2	2+	2+
Treatment arms	**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 and repetitions.	**1. Full treatment**: all farmers in a group got SMS. **2. Partial treatment**: half farmers in group got SMS. **Cross-randomized**: message framing and repetitions.
Second season text	Yes, rerandomized	Yes, all receive	Yes, rerandomized
Admin. outcome	1AF admin, LR2017 and LR2018	1AF admin, LR2018 and LR2019	1AF admin, 2018 and 2019
Coupon value	–	–	–
Baseline survey	No	No	No
Endline survey	Yes (phone), LR 2017	No	No

Notes: All experiments included a control group in addition to the treatment arms. SR and LR denote the short and the long rains agricultural seasons in Kenya, respectively. Topdressing is denoted as td. Treatment arms (#) denotes the number of treatment arms, and for 1AF "+" indicates that there were cross-randomizations in these samples both for the number of messages (1–5) and for framing.

farmers opted into the program. The baseline sample consisted of 832 farmers, of whom 733 completed an endline survey. From these, about two-thirds were females, and before treatments started, only 7 percent reported ever using lime, whereas 84 percent had used chemical fertilizers during the previous season (Table 1 and online Appendix Table C1).

IPA and PxD.—Innovations for Poverty Action (IPA) is a research and policy organization, and Precision Development (PxD) is a nonprofit organization supporting the provision of phone-based information services. We discuss the evaluation of two text message programs, IPA/PxD1-K and IPA/PxD2-K, implemented in Kenya through a partnership between these two organizations.

The IPA/PxD projects prioritized providing farmers with actionable advice. Recognizing that few smallholder farmers could afford conducting individual soil tests on their farms to assess their soil acidity, the programs used information from area-level soil tests to generate recommendations (online Appendix K describes soil data sources and how recommendations were built). Based on these soil samples, 82 percent of the IPA/PxD1-K sample and 77 percent of the IPA/PxD2-K sample were recommended to experiment with lime microdosing in a small area of their land.[8]

IPA/PxD1-K: The first program was implemented during the 2016 short rains season. Participants were identified using existing regional farmer databases. A sample of 1,897 farmers completed a phone-based baseline survey and were later randomized into two treatment arms or a control arm. Farmers in the first treatment arm (*General SMS*) received messages that mentioned the purpose of lime and recommended using fertilizers, but did not reference any soil test data. The messages received by farmers in the second arm (*Specific SMS*) referred to the area-level soil data and contained more precise details, such as additional guidance on recommended input quantities. Among those randomized into the treatment groups, 95 percent agreed to receive the messages. Depending on the intervention arm, between 24 and 28 messages were sent. Of these, 7 to 9 messages dealt with soil acidity and lime, 4 to 9 were about fertilizers, and the rest covered topics related to other agricultural practices. During the following agricultural season, the 2017 long rains, both treatment arms received five identical messages promoting the use of agricultural lime (in areas where lime was recommended). The control group remained untreated.

The sample for this experiment was 37 percent female. At baseline, only 18 percent had used one of the key recommended fertilizers in the previous season, and approximately 12 percent had ever used lime (Table 1 and online Appendix Table C2).

IPA/PxD2-K: A second program was implemented during the 2017 long rains season, targeting a different sample of farmers and reaching two additional areas.

[8] Agricultural lime is cheap but bulky. Farmers often find it challenging to transport and store it in sufficient quantities for widespread application. As a solution, farmers in the IPA/PxD programs were advised to opt for microdosing, targeting the base of the plants. This approach entails a lower dosage but requires reapplication each season. In experimental plots, lime microdosing increased yields by up to 14 percent (1AF 2014).

This program sent 13 messages solely focused on lime and fertilizers. Messages were sent right before the time for planting or top-dressing. Farmer recruitment was done via agricultural supply dealers (agrodealers), who invited existing clients to register to participate.[9]

Once registered by the agrodealer, farmers completed a brief phone baseline survey. They were then randomly assigned to one of three treatment arms or a control group. The first arm only received text messages (*SMS only*). The other two arms were designed to investigate whether real-time communication with a person could strengthen the texts. In the second arm (*SMS* + *Call*), farmers received a phone call from a field officer, and in the third arm (*SMS* + *Call Offer*) farmers could request to receive a call. The final sample consisted of 5,890 farmers, of whom 34 percent were female, 9 percent had used lime in the past, and 19 percent had used the key recommended fertilizer (Table 1 and online Appendix Table C3).

1AF.—One Acre Fund (1AF) is a social enterprise operating across six countries in Eastern and Southern Africa. 1AF's model relies on training farmers in modern agricultural techniques and providing them with inputs on credit early in the season, which they later repay. 1AF clients form groups of eight to eleven farmers and are supported by 1AF staff.

One of 1AF's products is agricultural lime. However, demand for lime was relatively low across their operating locations. Hypothesizing that this could reflect a lack of awareness, 1AF implemented two text message programs in Kenya and one in Rwanda to encourage lime use (1AF1-K, 1AF2-K, and 1AF3-R).[10] Messages were sent before the 1AF "enrollment season," that is, the period in which farmers register for the 1AF input program and place orders. To build area-level acidity recommendations, 1AF used its own soil data (see online Appendix K).

1AF1-K: 1AF's first text message intervention was implemented in western Kenya during the 2017 long rains season. Participants were randomly selected from lists of previous 1AF clients in Busia and Kakamega counties. The program sent six 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 mentioning the predicted level of acidity in the area, the amount of lime recommended and the expected returns to its application (*Detailed SMS*). A sample of 4,884 farmers participated in the experiment.

During the following long rains season, participants were rerandomized into treatment and control. The treatment group received messages promoting lime adoption.

[9] This method offered several advantages. First, it was a low-cost and quick method to recruit farmers. Second, farmers who are clients of agrodealers are already more likely to acquire inputs, and therefore might be able to benefit more from informational programs.

[10] Relative to farmers in other samples who rarely had contact with extension officers, 1AF 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 1AF field officers, who already followed lengthy farmer training protocols. In addition, all of their text message 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 percent of farmers using this service.

In total 2,931 farmers were rerandomized into receiving messages, allowing us to compare outcomes between three groups: farmers who were never assigned to receive messages, farmers who received messages during two consecutive seasons, and farmers who were only treated in the first season.

1AF2-K: The second program took place in Bungoma, Busia, Kakamega, and Vihiga counties during the 2018 long rains season, engaging 32,572 farmers recruited from previous 1AF client listings. Farmers were randomized at the individual level into a control group, a *Lime only* treatment group receiving lime-related messages or a *Lime* + *CAN* treatment group assigned to receive information on both lime and CAN fertilizer. The larger sample size allowed for cross-randomizing message content or framing (basic, highlighting yield increases, encouraging experimentation, making social comparisons, and promoting self-efficacy) and repetitions (one to five messages). The message framing was also cross-randomized to address the whole family instead of the individual. In the following season, all farmers, including the control group, received lime messages, effectively ending the lime experiment but allowing us to study effect persistence on CAN fertilizer use.

1AF3-R: The third program launched across Rwanda in 2018 and only sent lime-related texts. To identify spillover effects, the experiment was designed as a two-staged randomized experiment. 1AF farmer groups were randomly assigned into three arms: a full control arm (*Full Control*), where no farmers within a group received messages; a fully treated arm (*Full Treatment*), where all farmers with phones received messages; and the partially treated arm (*Partial Treatment*), where farmers with phones were further randomized into either receiving messages or remaining as controls. This design allows studying the extent of spillovers by comparing the outcomes of untargeted farmers in partially treated groups against those of farmers in the full control group. Additionally, we can estimate spillovers on individuals who do not own mobile phones by comparing non–phone owners in the treatment and the full control group.

The sample included 20,944 farmer groups, comprising 202,972 farmers. To study the direct effects of the program, we focus on the sample of 82,873 farmers who, at baseline, had phones and were not assigned to the control condition in the partially treated groups.

Content and repetition were also randomized among treated farmers. The content variations included a simple general message, as well as messages that highlighted yield impacts, encouraged self-diagnosis, referred to the use of soil tests, explained how lime works, encouraged farmers to order immediately, and highlighted soil acidity and yield impacts. The messages were further cross-randomized to be framed in terms of yield losses or gains.

The following season, farmer groups were rerandomized into treatment or control, with participation limited to those farmers who had enrolled in 1AF's 2018 input program. Farmers in treated groups were randomly assigned to receive or not receive messages. This introduced random variation to study effect persistence, and the impacts for those who were treated over two consecutive seasons.

C. *Considering Heterogeneity in Lime Requirements*

Mobile phones, unlike other mass distribution channels like radio or television, allow for personalized communication. This proved important in delivering information about agricultural lime, which is only recommended for acidic soils. IPA/PxD and 1AF capitalized on this by sending messages based on area-level soil pH information rather than providing blanket recommendations at the district or province level. The goal was to improve upon the prevailing situation, where farmers get no guidance or overly generic recommendations.

While the aim was to offer targeted information, the actual benefits of using lime depend on the specific soil conditions of each farm. For instance, certain farmers in regions classified as acidic might have been advised to experiment with lime, despite their individual plots not requiring it. Each organization recognized this issue and thus urged farmers to experiment or seek more information before making substantial lime investments. IPA/PxD recommended farmers to initially experiment on a small area of their farm. KALRO advised individual soil testing before lime application. Meanwhile, 1AF used its extensive on-the-ground extension network to offer additional assistance as needed.

However, to better understand the extent of heterogeneity in soil acidity within the treated areas, we analyze data from over 9,000 soil tests conducted in the areas of study. To the extent that the farms where these soil tests were conducted are representative of the land of participating farmers, we can roughly estimate the proportion of farmers likely to benefit from lime. A naive estimate of this share suggests that within areas where lime was recommended, 90 percent of soil tests were below a pH of 6.0, 68 percent were below a pH of 5.5, and almost all were below a pH of 7 (online Appendix Table K1, columns 2–4). Yet soil testing is prone to many sources of error. For instance, we found that the test-retest correlation of pH was 0.7 in a subset of soil samples blindly tested twice by the same soil laboratory (Fabregas et al. 2020b). Adjusting our estimates for this type of measurement error, we predict that over 96 percent of soil samples in the areas where experimentation with lime was recommended fell below the pH threshold of 6 (see online Appendix K for details).

In practice, using data from 1AF's lime experimental plots with corresponding soil tests, we find no significant difference in lime's impact on maize yields between farms with soil pH below or above 5.5 (online Appendix Figure K2). This also suggests that using strict soil acidity thresholds based on a single soil test for a land plot might be too restrictive.

We also assess the differential costs treated farmers might incur when using lime. The risk of overliming and creating soil alkalinity was minimal at the recommended quantities (see online Appendix K). Thus, the main costs would have been related to the time and money spent on acquiring and applying the input. Fortunately, lime is very cheap. We calculate the extra expenditure on lime between treatment and control farmers to be under a dollar, even conditioning on purchasing a positive amount of lime. Therefore, the incurred costs seem reasonable in relation to the learning value regarding lime's impact, especially considering the relatively high cost of individual soil tests ($15–20).

D. *Data*

Baseline Data.—For the KALRO sample, an in-person baseline survey was conducted before randomization, while phone-based baseline surveys were completed with farmers in the IPA/PxD1-K and IPA/PxD2-K programs. These surveys collected data on demographics, previous agricultural practices, and input use. For the 1AF projects, we rely on client administrative data from previous seasons, detailing gender and historical input purchases from 1AF.

Endline Data.—Survey data were collected for the KALRO, IPA/PxD1-K, IPA/PxD2-K, and 1AF1-K programs. Additionally, there is at least one administrative measure of real input acquisitions for each program. For KALRO and IPA/PxD, we use data from discount coupons that treatment and control farmers could redeem at local agrodealers. The coupons were devised as a way to observe real input choices while minimizing experimenter demand effects. While any discounts should have affected treatment and control group farmers in similar ways, this approach means that we observe demand at prices that would not have existed in the absence of these programs. For the 1AF projects, we measure input choices through the input orders placed with 1AF.[11]

KALRO: After the 2015 short rains season, farmers participated in an in-person endline survey, which contained knowledge and input use questions. During this visit, treatment and control group farmers received two paper coupons redeemable at selected agrodealers in their nearest market center. The first coupon offered a 50 percent discount on agricultural lime, and the second coupon offered a 50 percent discount on the farmer's choice of chemical fertilizer (CAN, DAP, NPK, or mavuno). Each coupon had a unique ID to trace redemption back to the respondent. Agrodealers were instructed to retain the coupons and record farmers' input purchases. For this sample, the survey questions measure input use concurrent with the program implementation, whereas the coupons measure input purchases corresponding to the subsequent agricultural season.

IPA/PxD: All treatment and control farmers in the IPA/PxD programs received input discount coupons via text early in the season. For the IPA/PxD1-K sample, both lime and fertilizer coupons were sent during the 2016 short rains. The lime coupon allowed farmers to opt for either 10 kg of lime or a bar of soap, the goal being to reveal farmers' true preferences without liquidity constraints. The fertilizer coupon offered a 30 percent discount on selected top-dressing fertilizers (mavuno, urea, or CAN). In the following season, the 2017 long rains, all farmers received digital lime coupons for a 15 percent discount. In total, 32 agrodealers in 25 distinct market centers participated in coupon redemption for this sample. A phone endline survey

[11] Given that the program contexts we investigate contained such diverse characteristics—which ranged from the extent of farmers' liquidity constraints and area-level cellphone signal strength, to local maize prices and regional availability of agrodealers—we take the heterogeneity in input prices simply as an additional source of variation.

was also conducted during the 2017 long rains. The survey included questions about input use during the 2016 short rains and 2017 long rains.

Farmers in the IPA/PxD2-K sample received two digital coupons, one for a 15 percent discount on lime and another one for a discount on the top-dressing fertilizer of their choice. Coupons were distributed during the 2017 long rains. The program collaborated with 102 agrodealers across 46 market centers to facilitate redemption. A phone endline was also completed with this sample, carried out in two randomly assigned phases. One-third of farmers were contacted at the end of the 2017 long rains, while the remaining were reached during the subsequent short rains. The latter group answered questions about their input use in both seasons.

1AF: Outcomes for all 1AF programs were measured through the input orders placed with the organization. However, the text message interventions took place before the period when farmers could join 1AF's input program for that season. Across 1AF interventions, between 60 to 76 percent of farmers who received text messages later enrolled to acquire inputs from 1AF. While we do not find any evidence of differential 1AF input program enrollment by treatment status (online Appendix Table C7, panels D–F, columns 3 and 8), we take a conservative approach and define our outcome variable as lime purchased through 1AF, without conditioning on whether farmers were enrolled in the 1AF program at the time of the experiment.[12] Additionally, a phone-based survey was conducted towards the end of the 2017 long rains, involving a random subsample of approximately 30 percent of the 1AF1-K farmers.

II. Empirical Strategy

A. *Estimating Individual Program Impacts*

Our primary outcomes are *followed lime* and *followed fertilizer* recommendations. The indicator variable *followed lime* takes the value one for farmers in the treatment and control groups if the farmer used lime and lime was recommended (or would have been recommended) or if the farmer did not use lime and lime was not recommended (or would not have been recommended).[13] *followed fertilizer* takes the value of one if farmers purchased at least one of the key recommended fertilizers for which administrative data is available and set to zero otherwise (key recommended fertilizers are listed in Table 1). For programs for which we have access to survey data, we can also measure changes in agricultural knowledge and use of other inputs.

[12] This will tend to underestimate impacts since farmers who did not enroll in the 1AF input program would not have been able to buy inputs from the organization. We show effects conditioning on enrollment in online Appendix Table D2, columns 5, 6, 11, and 12. We also look for any differential 1AF enrollment by treatment status over a second season (for the sample from which persistence effects are estimated) and, again, do not find statistically significant differences (online Appendix Table C7, panels D and F, columns 5 and 10).

[13] The 1AF programs recommended positive amounts of lime to all farmers. KALRO recommended farmers to test their soil and use lime 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.

We estimate intention-to-treat (ITT) effects.[14] The general equation we estimate for each program is

$$y_i = \alpha + \beta Treatment_i + \gamma_k + \epsilon_i, \tag{1}$$

where y_i is the outcome measure for farmer i. $Treatment_i$ denotes a dummy variable indicating treatment, γ_k is a vector of randomization strata fixed effects (if used in the specific experiment), and ϵ_i is the error term. Since some 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 online Appendix F and discuss lessons from these experimental variations in Section IV.

For binary outcomes, we estimate linear probability models and nonlinear analogs to equation 1 using a logistic regression model and report results in terms of odds ratios. This modeling choice reflects a concern to select an appropriate summary measure for a metanalysis, as discussed further below.

Our main specifications only control for the strata used in each randomization, as applicable. In the case of survey data, we also include enumerator-fixed effects. As a robustness check, the online Appendix shows results where we incorporate controls for other farmer demographic characteristics, baseline farming practices (as available), and location-fixed effects. Online Appendix B contains a list of strata and other controls used for each project.

For randomizations at the individual level, we do not cluster standard errors. For the 1AF3-R experiment, error terms are clustered at the farmer group level.

Validity of the Experimental Designs.—Online Appendix Tables C1–C6 show baseline characteristics by treatment status together with tests of equality of means across treatment arms for each program. Treatment and control arms are balanced along most characteristics. While some differences are statistically significant, *F*-tests of joint orthogonality for baseline variables that use specifications that include stratification variables fail to reject the null that coefficients are jointly zero in each experiment.[15]

Endline survey completion rates ranged from 79 percent (IPA/PxD1-K and 1AF1-K) to 92 percent (KALRO). We do not find any evidence of differential attrition by treatment status (online Appendix Table C7, panels A–D, columns 1 and 6). In the IPA/PxD and the 1AF1-K surveys, certain questions regarding lime use and knowledge were conditional on planting maize during the 2017 long rains season

[14] Some farmers might not have received or read the messages. The technology used to text farmers does not make it possible to determine whether the messages were opened. Using survey data, we document the following fraction of farmers who reported receiving agricultural information via text at endline: 54 percent in the 1AF1-K trial, 67 percent in the KALRO trial, 81 percent in the IPA/PxD2-K trial, and 92 percent in the IPA/PxD1-K trial. These differences could be explained by cellphone signal reception, the share of incorrect numbers held by each organization, the amount of time between receiving messages and completing the survey, and the way in which questions were asked. Since there is uncertainty around how precise these self-reports are, we do not attempt to scale the effects.

[15] We reject the null of joint orthogonality for one of the treatment arm comparisons in the 1AF2-K project, though we fail to reject it once we include controls and area-fixed effects. For all projects, we show *p*-values of these *F*-tests controlling for randomization strata and also when adding additional controls.

(corresponding to 97 percent, 98 percent, and 94 percent of those who completed the IPA/PxD1-K, IPA/PxD2-K and 1AF1-K surveys, respectively). To keep samples consistent across various outcomes, we condition all outcomes for those projects on maize cultivation during that season. Importantly, there are no significant differences in the probability of having differential missing data by treatment status with this additional condition (columns 2 and 7 in panels B to D).

B. *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 can vary across studies and are normally distributed. The weighted average effect, therefore, represents the mean of the distribution of true effects. Formally, the model can be written as

$$T_j = \mu + e_j + \zeta_j, \tag{2}$$

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, $e_j \sim N(0, \sigma_j^2)$ and $\zeta_j \sim N(0, \tau^2)$. σ_j^2 is the within-study variance for study j, while τ^2 is the between-study variance. The estimate of μ is

$$\hat{\mu} = \frac{\sum_{j=1}^{s} w_j T_j}{\sum_{j=1}^{s} w_j}, \tag{3}$$

where w_j are study-specific weights given by the inverse of the variance and s is the number of studies. In this case,

$$w_j = \frac{1}{\left(\hat{\tau}^2 + \hat{\sigma}_j^2\right)} \tag{4}$$

and in practice, we estimate τ^2 using the DerSimonian and Laird method (Der Simonian and Laird 1986). We conduct robustness checks using a number of alternative estimation methods (Sidik-Jonkman, restricted maximum likelihood, 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, which is the percentage of variability not explained by sampling error (Higgins et al. 2003; Higgins and Thompson 2002).[16]

[16] The Q-statistic is a χ^2 statistic with s minus 1 degrees of freedom and is calculated by:

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

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

We also estimate 95 percent prediction intervals.[17] For situations where there are multiple outcomes per study, we compute the mean of the effect sizes for each study and estimate standard errors accounting for within-trial correlations (Borenstein et al. 2017).[18]

Our preferred summary effect measure for binary outcomes is odds ratios, a relative effect measure. Using percentage point differences in meta-analyses has been empirically shown to lead to summary statistics that are less consistent than when using other metrics (Deeks 2002; Engels et al. 2000). Choosing a summary effect statistic that gives values that are similar for all studies also makes it more reasonable to express the effect as a single number. Nevertheless, in the results section, we discuss meta-analytic effects using both types of effect metrics for binary outcomes: odds ratios and percentage point differences.[19]

Finally, we complement this meta-analysis in two ways. First, by pooling all datasets together and estimating a single model (as in equation 1) including experiment dummies. Second, in online Appendix I, we present the meta-analysis results using Bayesian hierarchical random-effects models (Rubin 1981; Gelman et al. 1995).

III. Main Results

A summary of the main meta-analytic effects and accompanying tests of heterogeneity are reported in Table 3. We discuss them in this section.

---

by sampling error and is given by

$$I^2 = \max\left\{0, \frac{Q - (s-1)}{Q} \times 100\%\right\}.$$

I^2 is less sensitive to the number of studies included, but it depends on their precision (Borenstein et al. 2017). While there is subjectivity in interpreting the magnitudes, Higgins et al. (2003) provide 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 percent confidence interval.

[17] Prediction intervals provide a predictive range of future effects in exchangeable settings, accounting for uncertainty in the estimated effect, but also between-trial heterogeneity. They are estimated through the formula $\hat{\mu} \pm t \times \sqrt{\hat{\sigma}_\mu^2 + \hat{\tau}^2}$ where t denotes the critical value from a student's t distribution and $\hat{\sigma}_\mu$ the standard error of the weighted average.

[18] For each program, we compute the average effect by calculating the mean of the log-odds effects associated with that program. Subsequently, we calculate standard errors for these effects. To determine the within-study covariance matrix, we employ a bootstrap approach where we simulate 1,000 datasets. For each dataset, we assess the treatment effect on each outcome and then estimate the correlations between these effects (Bujkiewicz et al. 2019). Additional sensitivity analyses show that the results are robust to different assumptions about the correlation across outcomes, including both 0 and 1.

[19] When dealing with binary outcome variables, a potential issue with using the difference between two probabilities as the measure of choice is that the underlying baseline probabilities of the outcome in the population limit the range of variation for this difference. If values of the baseline probability of adoption between different studies vary, then the associated values of the difference in probabilities can also vary. This can give the appearance of heterogeneity in this measurement scale due to these constraints rather than due to other more substantial factors (Fleiss and Berlin 2009). Conceptually, the impacts of informational programs may depend on the baseline level of input adoption. When adoption is low, program effects might be smaller because it might be harder to persuade farmers to use the inputs. Once the technology is more common, farmers might become more responsive to the information. Finally, once a large share of farmers have adopted the technology, persuading the remaining farmers might be difficult. This S-shaped cumulative adoption pattern is often seen in models where there is heterogeneity among adopters and the distribution of values placed on the new technology by potential adopters is approximately normal (Hall and Khan 2003). In such cases, relative effect measures like odds ratios may be preferable to absolute measures like percentage point differences.

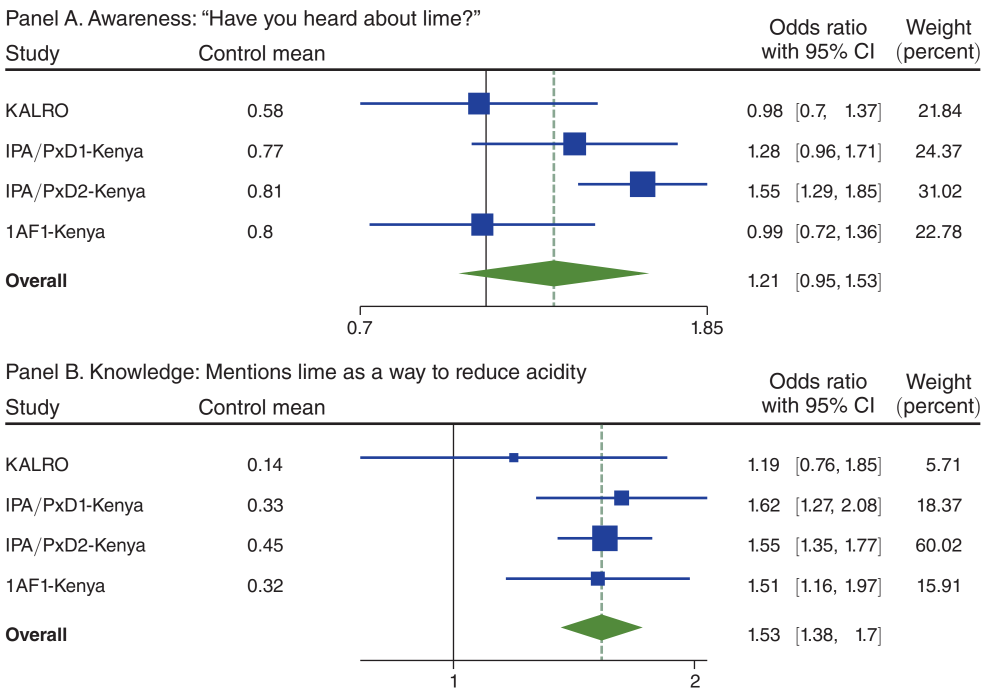

Figure 1. Effects on Lime Knowledge and Awareness

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 as odds ratios. The horizontal lines denote 95 percent confidence intervals.

A. *Impacts on Awareness and Knowledge*

We start by investigating impacts on the awareness and knowledge about agricultural lime. We focus on lime because it was a relatively unknown input, and encouraging its use was the main focus of all the programs.[20] Figure 1 shows that the effect for farmers having heard of lime (*awareness*), expressed as an odds ratio, is 1.21, but it is statistically insignificant (95 percent CI 0.95 to 1.53). However, there is substantial heterogeneity in this result. The p-value of the Q-statistic is 0.03, and $I^2 = 68\%$ (Table 3, panel A).

In contrast, the text messages increased the share of farmers who knew that lime is a remedy for soil acidity (*knowledge*). Across projects, this was recorded as free text without prompting and coded into categories. The meta-analytic odds ratio is 1.53 (95 percent CI 1.38 to 1.70). Moreover, we cannot reject the null of homogeneous treatment effects on knowledge. The p-value of the Q-statistic is 0.68 and the $I^2 = 0$ (95 percent CI 0 percent to 85 percent).

[20] Individual project results are shown in online Appendix Table D1. No equivalent questions were asked about recommended chemical fertilizers during the endline surveys across projects.

Table 3—Summary of Meta-analytic Results

Outcome	Observations (1)	Effect (2)	95% CI (3)	95% PI (4)	Q-stat (p-value) (5)	I^2 (%) (6)	I^2 95% CI (7)	τ^2 (8)
Panel A. Effects from logit specifications (odds ratios)								
Awareness (lime)	4	1.21	(0.95, 1.53)	(0.44, 3.29)	0.03	68	(6,89)	0.04
Knowledge (acidity)	4	1.53	(1.38, 1.70)	(1.21, 1.93)	0.68	0	(0,85)	0.00
Lime recommended	6	1.19	(1.11, 1.27)	(1.04, 1.36)	0.29	19	(0,63)	0.00
Fertilizer recommended	4	1.27	(1.15, 1.40)	(1.03, 1.57)	0.67	0	(0,85)	0.00
Recommended inputs	6	1.22	(1.16, 1.29)	(1.13, 1.32)	0.50	0	(0,75)	0.00
Other inputs	5	1.00	(0.93, 1.08)	(0.80, 1.25)	0.08	53	(0,83)	0.00
Persistence lime	4	1.06	(0.95, 1.18)	(0.84, 1.34)	0.71	0	(0,85)	0.00
Fatigue lime	3	1.29	(1.15, 1.45)	(0.61, 2.72)	0.82	0	(0,90)	0.00
Persistence fertilizer	4	1.08	(0.99, 1.19)	(0.88, 1.33)	0.60	0	(0,85)	0.00
Panel B. Effects from linear probability models (percentage points)								
Awareness (lime)	4	0.03	(−0.00, 0.06)	(−0.09, 0.14)	0.11	51	(0,84)	0.00
Knowledge (acidity)	4	0.08	(0.04, 0.12)	(−0.08, 0.24)	0.04	64	(0,88)	0.00
Lime recommended	6	0.02	(0.01, 0.03)	(−0.02, 0.06)	0.00	78	(51,90)	0.00
Fertilizer recommended	4	0.01	(0.00, 0.03)	(−0.04, 0.07)	0.02	70	(14,90)	0.00
Recommended inputs	6	0.02	(0.01, 0.03)	(−0.02, 0.05)	0.00	81	(60,91)	0.00
Other inputs	5	0.00	(−0.00, 0.01)	(−0.01, 0.01)	0.17	38	(0,77)	0.00
Peristence lime	4	0.00	(−0.00, 0.01)	(−0.01, 0.02)	0.79	0	(0,85)	0.00
Fatigue lime	3	0.02	(0.01, 0.02)	(−0.03, 0.07)	0.46	0	(0,90)	0.00
Peristence fertilizer	4	0.01	(−0.00, 0.02)	(−0.01, 0.03)	0.50	0	(0,85)	0.00
Panel C. Effects on quantities (Kg)								
Kg lime	6	1.18	(0.10, 2.27)	(−2.33, 4.70)	0.00	86	(72,93)	1.29
Kg fertilizer	4	0.43	(−0.03, 0.89)	(−1.37, 2.24)	0.02	70	(15,90)	0.12

Notes: Results for each meta-analysis are presented by row. Column 1 reports the number of experiments included in the meta-analysis. Columns 2 and 3 display results from random-effects meta-analyses and corresponding 95 percent confidence intervals (CI), respectively. Column 4 reports 95 percent prediction intervals (PI). Columns 5–8 provide information on heterogeneity measures. Panel A reports results in terms of odds odds ratios, estimated using logit. Panel B reports results estimated using linear probability models. Panel C reports results in kilograms.

When summarizing effects derived from linear probability models we find a significant increase of 8 percentage points in knowledge and an insignificant 3 percentage point effect on awareness (Table 3, panel B). Overall, we conclude that while farmers might have heard about this input regardless of treatment status, the programs were successful in conveying information about the purpose of this new technology.

B. *Impacts on Following Input Recommendations Using Administrative Data*

Next, we examine our primary outcomes and present our preferred estimates, which use administrative purchase data. This includes data concurrent with the implementation season for all programs, except for KALRO, for which we use

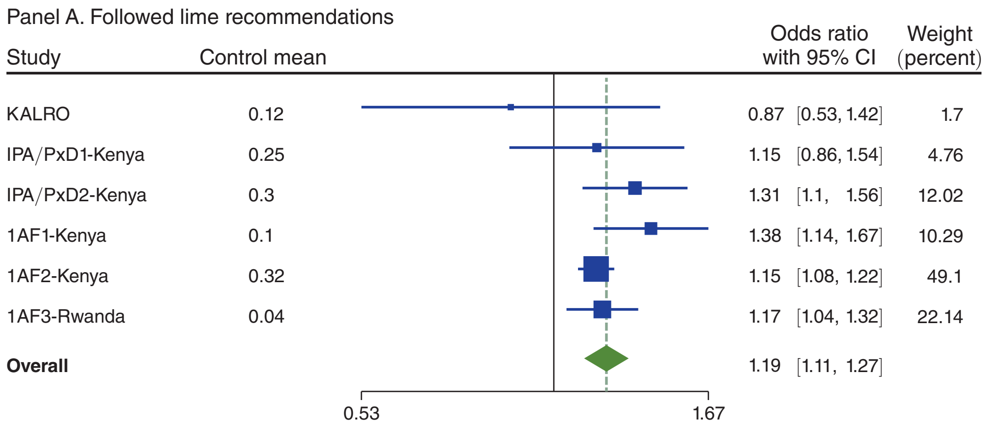

Figure 2. Effects on Lime Purchases (Administrative Data)

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 as odds ratios. The horizontal lines denote 95 percent confidence intervals. The KALRO results are measured using coupon redemption in the second season.

results based on coupon redemption for the subsequent agricultural season. In Section IVA, we discuss how estimated effects differ when using survey data.

Agricultural Lime.—We first examine impacts on the outcome all programs aimed to affect: following lime recommendations. Figure 2 shows that individual program effects range from a statistically insignificant odds ratio of 0.87 (95 percent CI 0.53 to 1.42) for KALRO to 1.38 (95 percent CI 1.14 to 1.67) for 1AF1-K.[21] The combined odds ratio for following the lime recommendation is 1.19 (95 percent CI 1.11 to 1.27).[22] The prediction interval, which gives a more intuitive sense of the range of effects of where a future sample would lie, ranges from 1.04 to 1.36. The Bayesian meta-analytic estimate is 1.20, and in that case, we estimate that 57 percent of observed heterogeneity is sampling variation (online Appendix Table I2).

Panel B in Table 3 shows the corresponding meta-analytic results estimated from linear probability models. The meta-analysis yields a combined effect of a 2 percentage point increase in the probability of following the recommendations (95 percent CI 0.01 to 0.03). In line with the discussion from Section IIB, using an absolute measure of effects, such as a percentage point difference, suggests a higher degree of true program heterogeneity. Indeed, we reject the null of homogeneous treatment effects across programs in this case. In this context, using odds ratios as a summary

[21] Figure 2 appears in the review piece by Fabregas, Kremer, and Schilbach (2019), which cites a working version of this paper.

[22] We also find reasonably consistent estimates for alternative specifications. Pooled data from all experiments into a single regression show that the odds of using lime increase by 14 percent or 1.3 percentage points (online Appendix Table E1, panel A, columns 1 and 3). The result is also robust to alternative methods of calculating τ^2 (online Appendix Table I1, panels B–D).

effect measure appears to better fit the data, as it entails a higher degree of effect consistency across studies.[23]

In terms of the unconditional quantity of lime acquired, the meta-analysis yields an estimate of 1.18 kgs purchased in areas where lime was recommended (online Appendix Figure I2). We reject the null of homogeneous effects for purchased quantities of lime, and estimate an I^2 of 86 percent (Table 3, panel C). This is perhaps not unexpected, given the variability in the amounts of lime recommended and the quantities that farmers could acquire across programs.

Fertilizers.—Only four programs (KALRO, IPA/PxD1-K, IPA/PxD2-K, and 1AF2-K) made fertilizer recommendations. Figure 3, panel A shows that the odds ratio for following the fertilizer recommendations is 1.27 (95 percent CI 1.15 to 1.40), and we fail to reject the null of homogeneous effects (Q-statistic p-value 0.67, $I^2 = 0\%$). The Bayesian results suggest a similar magnitude, 1.28, though the confidence intervals are wider (95 percent CI 0.86 to 1.82) (online Appendix Table I2).

Table 3, panel B shows effects using percentage point differences. We find a 1 percentage point increase in the purchase of recommended fertilizers, but we reject the null hypothesis of homogeneous treatment effects based on this meta-analysis. The impact on the unconditional amount of fertilizer acquired through vouchers or 1AF sales was 0.43 kg (95 percent CI −0.03 to 0.89) (online Appendix Figure I3).

The variable *followed fertilizer* 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. Therefore, we also estimate whether there was an overall increase in fertilizer purchases, for which we have administrative outcome data. The results are shown in Figure 3, panel B. The combined effect is 1.16 (95 percent CI 0.94 to 1.42). The smaller and statistically insignificant coefficient suggests some substitution between different types of fertilizers.[24] Altogether, the results are in line with the stated objectives of these programs: chemical fertilizers are well-known inputs, and messages shifted farmers towards experimenting with recommended blends. Moreover, we do not find evidence indicating that texts recommending an unfamiliar input (lime) had a greater effect than those recommending a familiar input (fertilizer).

[23] This is also in line with the notion that, beyond the heterogeneity arising from differences in baseline levels of input adoption, there is less impact heterogeneity coming from other project features. However, even when using our preferred odds ratio specification, a limitation of conducting a meta-analysis with only six studies is that the confidence intervals for I^2 tend to be quite large, making it difficult to be conclusive about the extent of program heterogeneity.

[24] The difference between Figure 3, panel A and Figure 3, panel B is driven by the PxD2-K program, in which different types of fertilizers were mentioned depending on local conditions. In Figure 3, panel B we code all fertilizers mentioned. Alternatively, online Appendix Figure I1 shows effects on all fertilizer purchases, using administrative data if it exists or survey data if it does not. In that case, we can also look at the effects of using any types of fertilizers, including planting fertilizers that are well-known in the region. The likelihood of purchasing any type of fertilizer is 1.14 (95 percent CI 0.97 to 1.33). Individual project results are shown in online Appendix Table D3.

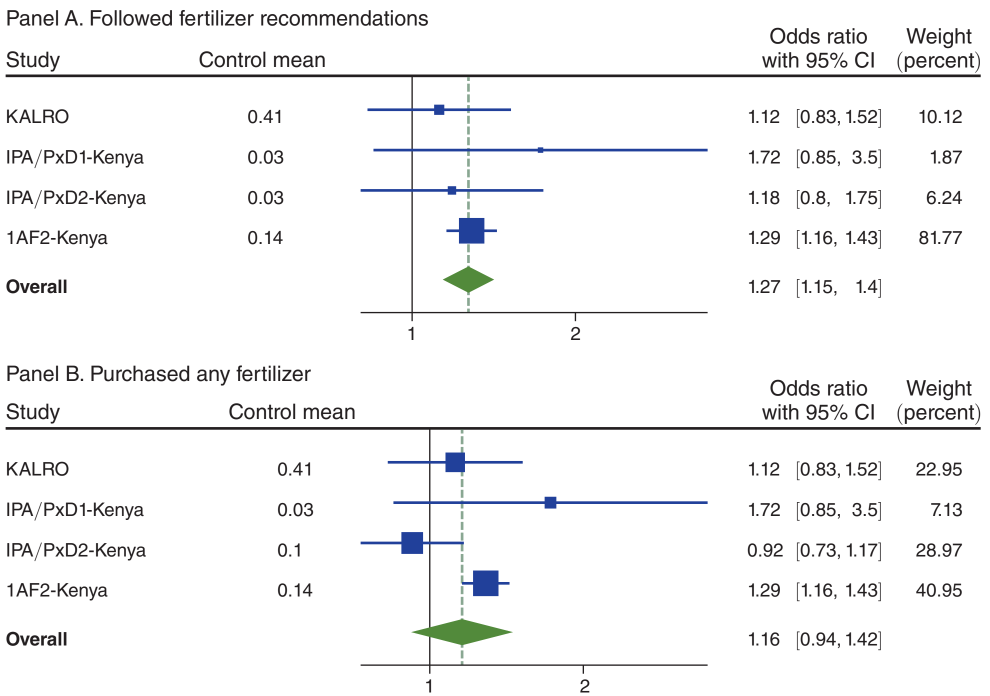

Figure 3. Effects on Fertilizer Purchases (Administrative Data)

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 as odds ratios. The horizontal lines denote 95 percent confidence intervals. The KALRO results are measured using coupon redemption in the second season. For panel B the dependent variable for IPA/PxD2-Kenya is a dummy equal to one if either urea or CAN were purchased.

C. *Combined Effects on All Recommended Inputs and Practices*

The previous section focused on lime and specific fertilizers as these were key program objectives, and we can measure outcomes through actual purchases. However, KALRO and IPA/PxD also provided information about other practices and inputs. In this section, we report overall effects, incorporating all potential adoption outcomes, using administrative data where possible and survey data otherwise. Online Appendix Table B1 reports the list of the inputs recommended and measured for each program.

To combine the effects on multiple outcomes, we follow two approaches. First, we incorporate multiple treatment effect estimates within studies, accounting for the fact that effects can be correlated within a study (Borenstein et al. 2009). Figure 4, panel A shows the corresponding forest plot. The estimated odds ratio is 1.22 (95 percent CI 1.16 to 1.29, $N = 6$), and we fail to reject the null of homogeneous treatment effects (p-value $= 0.50$). The prediction interval ranges from 1.13 to 1.32. The Bayesian estimate is 1.21, and under that model, we estimate that

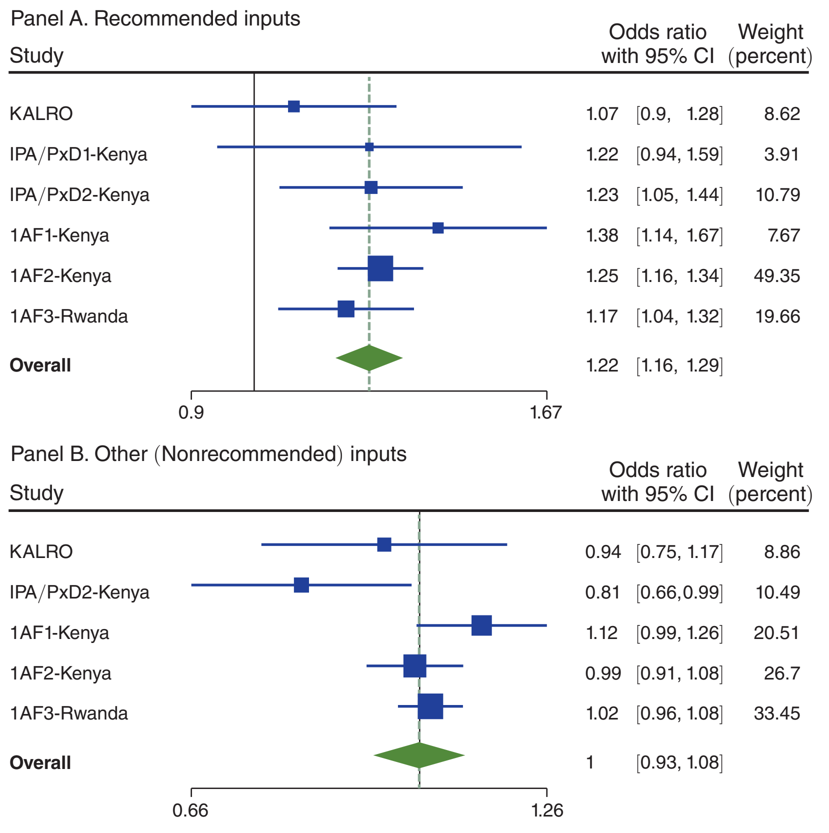

Figure 4. Effects on Recommended and Other (Nonrecommended) Inputs

Notes: The figure plots the meta-analysis results for the effect of the programs on the use or purchases of recommended inputs and other inputs not mentioned by the text messages. The effects are estimated using a random-effects meta-analysis model. Results are reported as odds ratios. The horizontal lines denote 95 percent confidence intervals. Panel A reports results for recommended inputs. Panel B reports results for other (nonrecommended) inputs. IPA/PxD1-K is not included in panel B because no data for other nonrecommended inputs was collected in that case.

62 percent of the observed heterogeneity is sampling variation (online Appendix Table I2).[25]

The meta-analytic results estimated from linear probability models suggest a combined effect of 2 percentage points (95 percent CI 0.01 to 0.03). Again, using an absolute effect measure suggests a higher degree of treatment heterogeneity (Table 3, panel B).

As a second strategy to combine outcomes, we standardize treatment effects for each experiment, following the construction of indices as per Kling, Liebman, and Katz (2007). Combining these point estimates through a meta-analysis, we find

[25] To study the impacts of 1AF2-K on all inputs we restrict the sample to the treatment arm that recommended both lime and CAN.

that the overall effect of the programs, expressed in standard deviations, is 0.06 (95 percent CI 0.03 to 0.08) (online Appendix Table I1, panel A).

We conclude that text messages appear to consistently affect farmers' input choices. Although the impacts are modest, they are not on a substantially different scale from those effects achieved through more intensive and costly extension approaches. To put these effect sizes into perspective, consider the effects of other agricultural informational interventions. Large in-person extension events in western Kenya increased the purchase of agricultural lime by 4 percentage points (Fabregas, Kremer, and Schilbach 2023). In India, Cole, and Fernando (2021) find that a more sophisticated voice-based service increased the adoption of recommended cotton seeds by 0.09 standard deviations, and BenYishay and Mobarak (2019) find increases of 3 to 7 percentage points in pit planting using in-person extension services through lead farmers. In Uganda, a video-based intervention increased the use of chemical fertilizers by 5 percentage points (Van Campenhout, Spielman, and Lecoutere 2021).

D. *Effects on Other* (*Nonrecommended*) *Inputs*

Does following the recommendations crowd out the purchase of other inputs? On average, we do not find that this is the case (online Appendix Table B1 lists all inputs we consider).[26] The combined effect on inputs that were not mentioned in the texts is negligible and statistically insignificant: 1.00 (95 percent CI 0.93 to 1.08, $N = 5$) (Figure 4, panel B).[27] However, we marginally fail to reject the null of homogeneity across results (p-value $= 0.08$, $I^2 = 53\%$) (Table 3, panel A).

E. *Effect Persistence*

To measure effect persistence, we measure input acquisition or use during a subsequent agricultural season ($t + 1$) for farmers that were only treated during the preceding season (t). Four experiments allow studying this question for lime: KALRO, IPA/PxD2-K, 1AF1-K, and 1AF3-R. We use administrative data for all of them except for IPA/PxD2-K, for which we only have survey data for the second season. For fertilizers, we exploit variation from: KALRO, IPA/PxD1-K, IPA/PxD2-K, and 1AF2-K, and use survey data for the IPA/PxD programs. Figure 5, panels A and B show results.

[26] We do not include nonrecommended fertilizers for programs that sent fertilizer recommendations, since farmers might have naturally substituted between different blends. These results are shown in online Appendix Figure I1. For the two programs that did not send any fertilizer recommendations (1AF1-K and 1AF3-R) we include fertilizer purchases. For 1AF2-K we only keep the treatment arm that recommended both lime and fertilizer to make the sample consistent with that of Figure 4 panel A. For the IPA/PxD1-K program only information for fertilizer and lime purchases was collected at endline, and therefore, the project is not included in this meta-analysis.

[27] Online Appendix Table D6 shows the results for recommended inputs, other inputs, and other nonrecommended fertilizers separately by experiment, using a seemingly unrelated regression framework to account for covariance across estimates. The results for the IPA/PxD1-K program suggest that farmers substituted 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 also negative for the IPA/PxD2-K program, although smaller and not statistically significant (panel C, columns 5 and 6). For that program, there is also evidence of input substitution in overall purchases (columns 3 and 4).

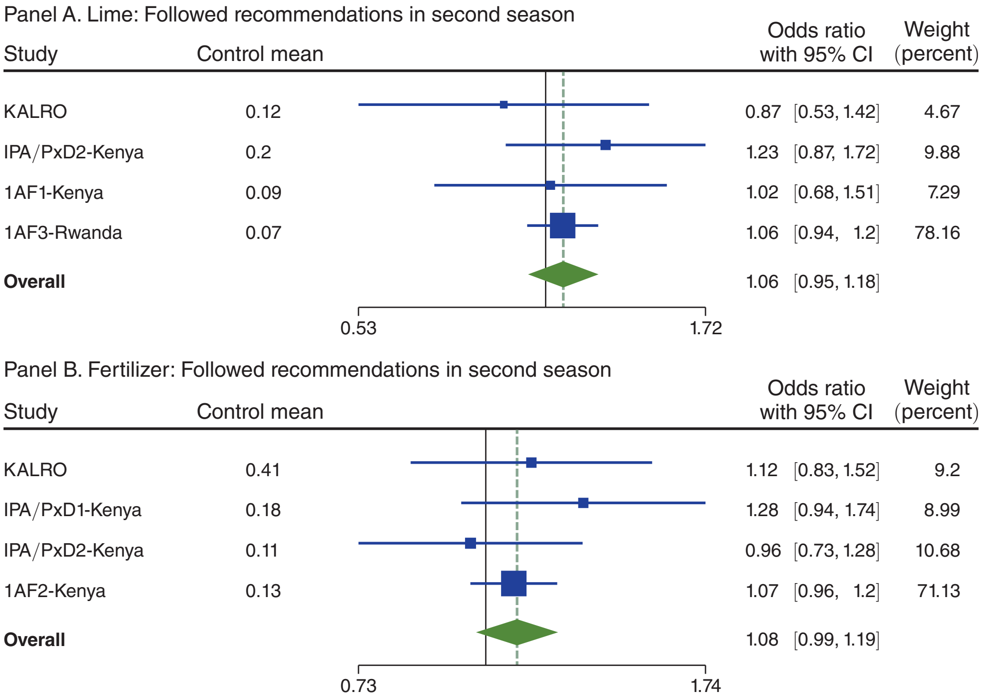

Figure 5. Effect Persistence over the Subsequent Season

Notes: The figure plots the meta-analysis results of following lime and fertilizer recommendations in the second season, for the subsamples that were only treated in the first season. The effects are estimated using a random-effects meta-analysis model. Results are reported as odds ratios. The horizontal lines denote 95 percent 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.

For individual projects, coefficients are mostly positive 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 and are, again, insignificant. For lime, the combined odds ratio is 1.06 (95 percent CI 0.95 to 1.18) and for fertilizer, 1.08 (95 percent CI 0.99 to 1.19). We fail to reject the null of homogenous effects in both cases (Table 3, panel A). 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.[28] Online Appendix Tables D4 and D5 show persistence results for each program using survey and administrative data separately.

[28] In the analysis of the 1AF3-R project, we include the second season's untreated farmers from partially treated groups. If we exclude them from the sample, we estimate a lime persistence odds ratio of 1.10 (95 percent CI 0.92 to 1.33) for that project, along with an overall meta-analytic effect of 1.09 (95 percent CI 0.94 to 1.26). Similarly, the corresponding fatigue effect for 1AF3-R is 1.29 (95 percent CI 1.08 to 1.54) with that restricted sample.

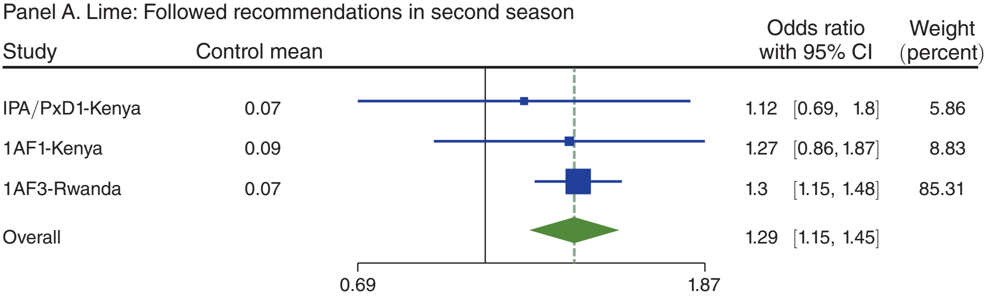

FIGURE 6. MESSAGE FATIGUE

Notes: The figure plots the meta-analysis results of following lime recommendations in the second season, for the subsample of farmers treated in both seasons, compared to the control group. The effects are estimated using a random-effects meta-analysis model. Results are reported as odds ratios. The horizontal lines denote 95 percent confidence intervals.

F. *Message Fatigue*

We also ask whether re-treating farmers during a second season sustained effects. To answer this question we look at lime purchases for the three programs that re-treated farmers during a subsequent season: IPA/PxD1-K, 1AF1-K, and 1AF3-R. We use administrative data in all cases. The combined effect in odds ratios is 1.29 (95 percent CI 1.15 to 1.45) (Figure 6) and we fail to reject the null of homogeneous treatment effects (p-value $=$ 0.8). The corresponding meta-analytic effect for those three programs measured over the first season is 1.22 (95 percent CI 1.11 to 1.35). We find a corresponding effect of 2 percentage points when using estimates from linear probability specifications (Table 3, panel B).

These results provide little empirical support for the idea that re-treating farmers with text messages will necessarily lead to fatigue. Moreover, we note that, under at least the simplest Bayesian models of learning, repeating messages would have no effect.

G. *Who Is Most Responsive to These Programs?*

A potential concern about digital-based approaches is that they will favor younger, more educated, or richer farmers. However, we find little evidence of heterogeneous effects by gender, age, level of education, or farm size (for 1AF this was proxied using the size of the input package bought). Heterogeneity results for each program are shown in online Appendix Tables H1–H2. To increase power, we also show results pooling all datasets together in online Appendix Table E2. Again, we find no evidence of a statistically significant differential program effect by these characteristics. Moreover, we find that input purchases were not differentially affected by whether farmers had used or heard about fertilizer or lime in the past. We interpret this finding as suggesting that the effects operate through channels other than simply raising knowledge about these inputs.

IV. Lessons from Individual Experiments

This section uses individual projects' experimental variation to gather lessons about impact differences by data sources, variation in message content, repetition, complementary services, and spillovers.

A. *Differential Effects of Self-Reported versus Administrative Data*

Are effects measured using self-reported versus administrative data equivalent? For lime we can compare the results of four projects for which we have both types of data (KALRO, IPA/PxD1-K, IPA/PxD2-K, 1AF1-K).

Online Appendix Figure I4 shows the meta-analysis of the ratio between the odds ratio coefficients obtained using survey data and those obtained using administrative records for lime.[29] The meta-analytic estimate for the difference in survey and administrative data is 1.19 (95 percent CI 1.03 to 1.38) indicating that the effects estimated using survey data are higher than those using administrative records.

Which programs drive this difference? A likely candidate is KALRO, since the administrative and survey data correspond to two different seasons. However, the difference between data sources is relatively small (online Appendix Table D2). Similarly, the survey data lines up reasonably well with the administrative reports for 1AF1-K. However, for the IPA/PxD programs, the survey results are statistically larger than the ones estimated using data from coupon redemption. This discrepancy could suggest either of two possibilities. Firstly, it is possible that the survey data is affected by social desirability or recall bias and that true lime purchases or use are misreported in the questionnaire. This could be the case, for instance, if farmers felt social pressure to report that they followed the recommendations even when they did not. A second possibility is that the coupon redemption underestimates true lime use if treated farmers obtained inputs from sources not captured by the administrative data.

We explore these possibilities for farmers in the IPA/PxD2-K sample, for whom we have more information. First, we examine whether farmers who potentially had access to alternative sources of lime due to their participation in 1AF programs are more likely to report they used lime even if they did not redeem the coupon. We find that participating in 1AF programs (36 percent of the sample) is associated with a 5 percentage point increase in the likelihood of following the recommendations in the survey but not in the administrative data (from 8 to 13 percent). This could suggest that some farmers might have procured lime from alternative sources.

We can also compare farmers' reports about which shops they acquired inputs from against data from a survey completed with agrodealers about their lime stock. Among the farmers who reported in the survey that they had acquired lime from an agrodealer (87 percent), those who reported using lime but did not redeem the lime coupon ("mismatched") were 5 percentage points more likely than those without a mismatch to be unable to report the shop from which they bought it. Conditional on listing an agrodealer, these mismatched cases were also 8 percentage points more

[29] A ratio of odds ratios compares 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.

likely to claim they acquired lime from a shop that, according to our monitoring data, had not stocked lime during that period (an increase from 9 to 17 percent). This hints at differential misreporting in the survey due to social desirability bias. Overall, true effects on lime use are likely to be between these two bounds, but we take the more conservative administrative data results as our preferred estimates.

When considering discrepancies in fertilizer use, we have data for three programs (KALRO, IPA/PxD1-K, and IPA/PxD2-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 percent versus 41 percent for KALRO, 15 percent versus 2 percent for IPA/PxD1-K, and 16 percent versus 2 percent for IPA/PxD2-K (online Appendix Table D3)). This likely suggests that farmers procured fertilizer from sources other than the participating agrodealers that redeemed coupons. However, relative to the discrepancy for lime, the direction of the gap between survey and administrative data impacts is negative and statistically insignificant.

B. *Effects of Message Framing, Content, and Repetition*

This section discusses the differential treatment arm effects in PxD/IPA1-K, PxD/IPA2-K, 1AF2-K, and 1AF3-R. The latter two cross-randomized message framings and repetition, and so the effects of treatments described should be interpreted as conditional on the distribution of the other treatments (Muralidharan, Romero, and Wüthrich, forthcoming).

Framing and Content.—Behavioral economics posits that the way in which information is presented can influence individual choices. If so, small adjustments in message framing or content could be an inexpensive way to improve the effectiveness of these programs. 1AF2-K and 1AF3-R randomized different versions of the input messages, with the intention of appealing to well-known behavioral biases or providing additional information to farmers. All 1AF2-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"). 1AF3-R messages were also cross-randomized to be framed either as a loss or a gain (e.g., "to increase yields" versus "to avoid a yield loss").

Online Appendix Table F2, panels A–B show effects for lime and fertilizer for 1AF2-K, panel C shows effects for lime for 1AF3-R. In columns 1–2 and 7–8 the reference group is the control, whereas in columns 3–4 and 9–10 the reference group are those assigned to the basic message. All versions appear to be equally effective. The only marginally statistically significant effect we detect is for messages that included information on the potential increase in yields for 1AF2-K, but the effect does not hold for fertilizer purchases.

In terms of framing, addressing the entire family was less effective for lime purchases (panel A, columns 5–6 and 11–12). The effects for fertilizer are also negative, but not statistically significant (panel B). For 1AF3-R we do not find consistent evidence of differential impacts with a loss framing (panel C, column 5–6 and 11–12).

Overall, we conclude that the way specific messages are presented had limited influence on whether farmers followed the recommendations. However, since the cost of optimizing messages is very low, this is an area that warrants further exploration.

Information Specificity.—Evidence on the role of information specificity comes from the IPA/PxD1-K and 1AF1-K projects. The IPA/PxD1-K project randomized farmers to either a general information arm or a treatment arm that provided specific information about the extent of soil acidity in the local area. While the treatment arm with specific information was significantly more likely to increase knowledge about lime, we do not find significant differences between the arms on the probability of purchasing the input (online Appendix Table F1, panel A). These results are similar to the effects of the *Broad SMS* and *Detailed SMS* treatment arms implemented by the 1AF1-K program. The point estimates between treatment arms are similar, and we cannot reject equality (online Appendix Table F1, panel C). We conclude that providing additional details on local soil characteristics made little difference in whether farmers followed the recommendations.

Message Repetition.—Message repetition has been identified in the communication and marketing literatures as a driver of consumer choices (Schmidt and Eisend 2015). 1AF2-K and 1AF3-R cross-randomized the number of repeated messages, where each message was sent every couple of days. We find evidence that repetition had modest positive effects on input purchases. The odds ratio effect of one additional text on purchasing lime is 1.03 in the 1AF2-K program (online Appendix Table F3, panel A, column 7) and 1.05 in the 1AF3-R program (panel C, column 7). This corresponds to a significant increase of 0.2 to 0.4 percentage points. For both programs, the effect is driven by receiving at least two messages, and we find no significant effects from receiving additional messages. For fertilizers, the odds increase by 7 percent or 0.7 percentage points (panel B).

C. *Are Text Messages Strengthened by Phone Calls?*

Text messages can be cheap, timely, and farmers can also consult them at later times as needed. However, texts are a restrictive and passive medium to convey information. To address these concerns, the PxD/IPA2-K project experimented with three treatment arms. In the first arm, farmers received only text messages (*SMS*). In the second arm, farmers received text messages and a phone call from a field officer (*SMS* + *Call*). In the third arm, farmers received the text messages and were offered the possibility of texting back to receive a call (*SMS* + *Call Offer*). The calls did not provide any additional information apart from what was contained in the messages, but farmers could ask clarifying questions. All calls were free.

We do not find high demand for an additional phone call. Only 13 percent of farmers assigned to the *SMS* + *Call Offer* group requested a call. This relatively low demand is in line with that of the 1AF projects, where a hotline was also available for all treated farmers, but where less than 1 percent 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 the lime recommendations (online Appendix Table F1, panel B).

D. *Are There Spillovers?*

These programs could create spillovers if beneficiaries share information with nonparticipants, who might also adopt the recommended technologies. If non-treated farmers benefit from the text message programs, we also risk underestimating overall impacts.

To assess the potential magnitude of these spillovers, we focus on the 1AF programs, since we can exploit the existence of untreated peers in farmer groups. However, we note that any potential spillovers might have been higher in this context compared to the IPA/PxD and KALRO programs where subjects operated individually.

We pursue three approaches. First, we use the variation created in the number of treated farmers within 1AF farmer groups. We estimate regressions of the following general form only for control farmers:

$$y_i = \alpha + \beta_1 \mathit{Treat\_Peers}_i + \beta_3 \mathit{Group\_size}_i + \gamma_k + \epsilon_i, \tag{5}$$

where we include the number of group members who are assigned to the treatment ($\mathit{Treat\_Peers}_i$) and include controls for group size. In this case, β_1 compares control group respondents who are exposed to a higher fraction of treated farmers. Online Appendix Table D8, columns 1–2 and 9–10 show results for 1AF1-K, 1AF2-K and 1AF3-R. We find no evidence of spillovers using this approach.

Second, to obtain cleaner evidence of spillovers, we leverage the two-staged 1AF3-R randomization. We compare untargeted farmers in partly treated groups against those in the pure control groups (online Appendix Table D8, panel D, columns 3–4 and 11–12). There is a statistically significant 14 percent increase in the odds of lime purchases for those who were untreated in the partly treated groups relative to the pure control group once all controls are included (or a 0.4 percentage point increase using a linear probability model). Columns 5–6 and 13–14 explore whether farmers without registered phones in the treated groups in 1AF3-R were more likely to adopt inputs relative to farmers without registered phones in non-treated groups. We find marginal statistically significant spillovers to these individuals once we include additional controls, with a 17 percent increase in the odds of following the recommendations among those without phones (a 0.3 percentage point increase).

Third, we estimate equation (5) for the population of non–phone holders in 1AF3-R (phone ownership is almost universal in Kenya, so we cannot use this approach for the other projects) and find some evidence of spillover effects, though not consistently statistically significant for different specifications (columns 7–8 and 15–16).

While we do not find consistent evidence of spillover effects across all three 1AF programs and different specifications, the evidence from the cleanest randomization design, within a program where farmers often interact, is suggestive of some spillovers for farmers in partially treated groups and for those without phones.

V. Cost-Effectiveness and Cost-Benefit Analysis

We present two types of calculations to assess program returns. First, we estimate the cost-effectiveness of a representative text-based program in comparison to non-digital approaches with similar goals. While this comparison alone is not dispositive for program investment decisions, it helps in evaluating extension approaches when taking policy goals as given.

Second, we conduct back-of-the-envelope calculations to estimate the marginal benefit-cost ratio of a representative text message intervention. While this calculation excludes the fixed costs associated with developing, managing and piloting messages, for large enough programs the overall benefit-cost ratio will be close to the marginal benefit-cost ratio. To determine the benefits of reaching an additional farmer, we combine information from the effects on lime and fertilizer use with existing agronomic data to roughly estimate impacts on yields and agricultural profits.[30]

For cost estimates, we use the marginal costs of the text messages (we also use marginal costs for the other in-person programs when making comparisons). The cost of sending one text message for the medium-sized programs was approximately \$0.01. These costs can be significantly lowered to \$0.001 if the programs operate at scale with bulk texting. For both calculations, we only focus on the use of inputs for a single agricultural season. Online Appendix G provides additional details.

Cost-Effectiveness.—Consider a program that sends three lime-related messages. The marginal cost per farmer would be \$0.003–\$0.03 per season. Using the most conservative marginal cost for this program, we estimate that the cost of inducing one farmer to experiment with lime is approximately US\$1.50. Using the summary effects from the quantity meta-analysis, we estimate that the cost per 10 kgs of lime used due to this type of program is \$0.25.

We contrast these estimates to those of in-person extension approaches implemented in the region. First, we compare them to those of Farmer Field Days (FFDs), an intervention implemented in western Kenya by KALRO (Fabregas, Kremer, and Schilbach 2023). The FFDs consisted of large in-person meetings with farmers where they could observe test plots and learn more about various inputs and practices, including agricultural lime. We estimate that the cost per farmer attended was \$9. The odds ratio of FFDs on lime purchases was 1.54 or a 3.8 percentage point increase in lime use, using a linear probability model. The point estimate on the quantity of lime purchased was an increase of 6.2 kg. Depending on whether we use odds ratios or percentage points, this translates to a per-farmer experimentation marginal cost of \$38–\$46, even when we only attribute a fifth of the overall FFD costs to lime teaching. The estimated cost per 10 kgs of lime purchased was \$2.8, more than ten times the estimated marginal cost from the text message program.

A second experiment conducted by 1AF in western Kenya tested lime sales incentives for field officers. These incentives were found to increase the probability of purchasing lime by 13 percentage points and the quantity of lime purchased by 6.6 kgs. The cost of this program involved a payment to field officers of \$0.5 per

[30] Experiments directly measuring the impact of text messages on farm profits would have the advantage of measuring the local average treatment effect (LATE) for those farmers who change behavior in response to the messages. However, given the per farmer cost of collecting profit data, the noise in measuring profits, and the small effects that would render a program cost-effective, studies powered to pick up cost-effective impacts would likely be prohibitively expensive.

adopting farmer, plus a day of training for the field officers (1AF 2019d). We estimate that the cost per experimenting farmer was \$1.88, while the cost per 10 kgs of lime purchased was \$0.38.

Therefore, text messages compare favorably to these in-person interventions over a single season, especially with bulk texting. However, a complete comparison across programs would also need to account for potential differences in the extent of spillovers and effect persistence, something we cannot assess for these comparison programs.

Cost-Benefit.—To approximate the lime benefits in terms of yields, we use the median of four agronomic trials in the region and calculate a 10.3 kg maize yield increase per 10 kg of lime applied (see online Appendix G). Since these experimental plots were, for the most part, implemented in regions deemed acidic but in farms with various levels of pH, the estimates of returns to lime already account for the fact that not every farm might have experienced an increase in yields.[31] We estimate that the profits obtained from an additional 10 kg of lime are approximately \$2.1, which takes into account the revenue from additional maize sales using prevailing market prices, minus the costs of applying lime and the additional labor costs from harvesting and transport. At the estimated lime application rate, we calculate a benefit-cost ratio of 8:1 for a three-message program. With at-scale unit costs of \$0.001 per text message, the implied marginal benefit-cost ratio is closer to 83:1.

For fertilizer, we use the impact of 10 kg of application on yields, 24.8 kg, from (Duflo, Kremer, and Robinson 2011). The cost of applying 10 additional kg of fertilizer is estimated to be approximately \$7.4, which considers the local price of fertilizer, transport, and application costs. Considering the overall impact of the programs in terms of the quantity of fertilizer applied implies a profit of \$0.07 per treated farmer. Considering a per-farmer program cost of \$0.04 (four-message program) the benefit-cost ratio is 1.75. However, at scale, with a unit cost of \$0.001 per text, the implied marginal benefit-cost ratio would be closer to 18:1. Combining the two components, lime and fertilizer, in a seven-message intervention, we obtain marginal benefit-cost ratios of up to 46:1 when programs are operated at a very large scale. [32]

These calculations should be interpreted with caution since they rely on many assumptions. However, they are encouraging for a number of reasons. First, the estimates on impacts are likely lower bounds. As discussed, there is suggestive evidence of information sharing among farmers, which we do not include. 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.

[31] In online Appendix K we show that even under very conservative assumptions about the fraction of farmers that might have benefited from these programs, the marginal benefit-to-cost ratio remains at 10 to 1 or higher when operating these programs at scale.

[32] Our finding that the absolute value of benefits per farmer is modest but still dwarfs marginal costs implies that some minimum number of farmers will be required to cover the fixed costs of designing and operating such programs but that asymptotically as the number of farmers served becomes large enough the overall benefit-cost ratio will approach the marginal benefit-cost ratio. Governments and telecommunication companies often have telephone numbers for and the ability to reach very large numbers of farmers, but some other organizations do not. We also discuss calculations with fixed costs in online Appendix G.

VI. Conclusion

An extensive body of literature in economics has identified informational barriers as a constraint to behavior change and technology adoption. The rapid uptake of cell phones 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 extend to different populations and contexts is critical for policy design. Some see the success of these programs as highly dependent on context and design details.

We experimentally evaluate 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. Although the programs were all run by well-known and trusted organizations, their target audiences, message designs, and specific content differed. While we cannot make conclusive statements about impact heterogeneity with only six projects, we failed to find strong evidence to support the idea that differences in target population, or the exact details of messages substantially affected the impacts of these programs.

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 nonsignificant results as evidence of no impact.

Our back-of-the-envelope calculations further suggest that although the effects we estimate are modest in absolute terms, text-based approaches are highly cost-effective from the point of view of an organization that is interested in promoting new inputs.

While we cannot fully disentangle the mechanisms through which these programs operate, we show that impacts decayed over time but re-treating farmers sustained effects. This may suggest 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 who knew nothing about the new technologies at baseline. Moreover, providing farmers with richer information and adding a phone call did not significantly change their behavior.

As more sophisticated technologies, such as smartphones and large language models, continue to advance and be adopted over time, opportunities for digital information provision are likely to improve. There is a large scope for policymakers and researchers to continue exploring how to effectively deliver information at scale in cheaper ways.

REFERENCES

1AF. 2014. *Managing Soil Acidity with Lime* (*2014 Phase 2 Trial Report*). Kakamega, Kenya: One Acre Fund. https://oneacrefund.org/documents/70/Managing SoilAcidityTrialReportAgInnovations.pdf. (accessed January 2018).

1AF. 2016. *Administrative Farmer Data for Kenyan Clients, 2016* (*Database*). Kakamega, Kenya: One Acre Fund (accessed May 2018).

1AF. 2017. *Administrative Farmer Data for Kenyan Clients, 2017* (*Database*). Kakamega, Kenya: One Acre Fund (accessed May 2018).

1AF. 2018. *Administrative Farmer Data for Rwandan Clients, 2018* (*Database*). Kakamega, Kenya: One Acre Fund (accessed June 2019).

1AF. 2019a. *Administrative Farmer Data for Kenyan Clients, 2018* (*Database*). Kakamega, Kenya: One Acre Fund (accessed June 2019).

1AF. 2019b. *Administrative Farmer Data for Kenyan Clients, 2019* (*Database*). Kakamega, Kenya: One Acre Fund (accessed June 2019).

1AF. 2019c. *Administrative Farmer Data for Rwandan Clients, 2019* (*Database*). Kakamega, Kenya: One Acre Fund (accessed on May 2020).

1AF. 2019d. "Kenya Lime Experience Trial Results, 2019."

1AF. 2020. *Lime Experience Trials, Kenya* (*Database*). Kakamega, Kenya: One Acre Fund (accessed May 2020).

1AF. n.d.a. *Soil Data for Rwanda, Rwanda* (*Database*). Kakamega, Kenya: One Acre Fund (accessed June 2019).

1AF. n.d.b. *Soil Data for Western Kenya, Kenya* (*Database*). Kakamega, Kenya: One Acre Fund (accessed September 2017).

Aker, Jenny C. 2017. "Using Digital Technology for Public Service Provision in Developing Countries." In *Digital Revolutions in Public Finance*, edited by Sanjeev Gupta, Michael Keen, Alpa Shah, and Geneviéve Verdier, 201–24. Washington, DC: International Monetary Fund.

Aker, Jenny C., and Marcel Fafchamps. 2015. "Mobile Phone Coverage and Producer Markets: Evidence from West Africa." *World Bank Economic Review* 29 (2): 262–92.

Aker, Jenny C., and Isaac M. Mbiti. 2010. "Mobile Phones and Economic Development in Africa." *Journal of Economic Perspectives* 24 (3): 207–32.

Aker, Jenny C., Ishita Ghosh, and Jenna Burrell. 2016. "The Promise (and Pitfalls) of ICT for Agriculture Initiatives." *Agricultural Economics* 47 (S1): 35–48.

Aker, Jenny C., Christopher Ksoll, and Travis J. Lybbert. 2012. "Can Mobile Phones Improve Learning? Evidence from a Field Experiment in Niger." *American Economic Journal: Applied Economics* 4 (4): 94–120.

Allcott, Hunt. 2015. "Site Selection Bias in Program Evaluation." *Quarterly Journal of Economics* 130 (3): 1117–65.

Angrist, Noam, Peter Bergman, and Moitshepi Matsheng. 2022. "Experimental Evidence on Learning Using Low-Tech When School Is out." *Nature Human Behaviour* 6 (7): 941–50.

Arouna, Aminou, Jeffrey D. Michler, Wilfried G. Yergo, and Kazuki Saito. 2021. "One Size Fits All? Experimental Evidence on the Digital Delivery of Personalized Extension Advice in Nigeria." *American Journal of Agricultural Economics* 103 (2): 596–619.

Baumüller, Heike. 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): 134–54.

Beg, Sabrin, Mahnaz Islam, and Khandker Wahedur Rahman. 2024. "Information and Behavior: Evidence from Fertilizer Quantity Recommendations in Bangladesh." *Journal of Development Economics* 166: 103195.

Benartzi, Shlomo, John Beshears, Katherine L. Milkman, Cass R. Sunstein, Richard H. Thaler, Maya Shankar, Will Tucker-Ray, William J. Congdon, and Steven Galing. 2017. "Should Governments Invest More in Nudging?" *Psychological Science* 28 (8): 1041–55.

BenYishay, Ariel, and A. Mushfiq Mobarak. 2019. "Social Learning and Incentives for Experimentation and Communication." *Review of Economic Studies* 86 (3): 976–1009.

Berazneva, Julia, Linden McBride, Megan Sheahan, and David Guerena. 2016. "Perceived, Measured, and Estimated Soil Fertility in East Africa: Implications for Farmers and Researchers." In *2016 Agricultural and Applied Economics Association Annual Meeting*. Boston, MA: Agricultural and Applied Economics Association. 10.22004/ag.econ.235466.

Bettinger, Eric, Nina Cunha, Guilherme Lichand, and Ricardo Madeira. 2021. "Are the Effects of Informational Interventions Driven by Salience?" University of Zurich Economics Working Paper.

Bird, Kelli A., Benjamin L. Castleman, Jeffrey T. Denning, Joshua Goodman, Cait Lamberton, and Kelly Ochs Rosinger. 2019. "Nudging at Scale: Experimental Evidence from FAFSA Completion Campaigns." *Journal of Economic Behavior and Organization* 183: 105–28.

Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Rothstein H. Rothstein. 2009. *Introduction to Meta-analysis*. New York, NY: John Wiley and Sons.

Borenstein, Michael, Julian P. T. Higgins, Larry V. Hedges, and Hannah R. Rothstein. 2017. "Basics of Meta-analysis: I2 Is Not an Absolute Measure of Heterogeneity." *Research Synthesis Methods* 8 (1): 5–18.

Brady, Nyle C., and Ray R. Weil. 2004. "Soil Acidity, Alkalinity, and Salinity." In *Elements of the Nature and Properties of Soils*, 266–315. London, UK: Pearson.

Bujkiewicz, Sylwia, Felix Achana, Tasos Papanikos, Richard D. Riley, and Keith R. Abrams. 2019. "NICE DSU Technical Support Document 20: Multivariate Meta-analysis of Summary Data for Combining Treatment Effects on Correlated Outcomes and Evaluating Surrogate Endpoints." University of Sheffield. https://www.sheffield.ac.uk/nice-dsu/tsds/full-list (accessed July 2022).

Buntaine, Mark T., Ryan Jablonski, Daniel L. Nielson, and Paula M. Pickering. 2018. "SMS Texts on Corruption Help Ugandan Voters Hold Elected Councillors Accountable at the Polls." *PNAS* 115 (26): 6668–73.

Camacho, Adriana, and Emily Conover. 2019. "The Impact of Receiving SMS Price and Weather Information on Small Scale Farmers in Colombia." *World Development* 123: 104596.

Casaburi, Lorenzo, Michael Kremer, Sendhil Mullainathan, and Ravindra Ramrattan. 2014. "Harnessing ICT to Increase Agricultural Production: Evidence from Kenya." Unpublished.

Chandrasekhar, Arun G., Esther Duflo, Michael Kremer, João F. Pugliese, Jonathan Robinson, and Frank Schilbach. 2022. "Blue Spoons: Sparking Communication about Appropriate Technology Use." NBER Working Paper 30423.

Chuang, Erica, Pascaline Dupas, Elise Huillery, and Juliette Seban. 2020. "Sex, Lies, and Measurement: Consistency Tests for Indirect Response Survey Methods." *Journal of Development Economics* 148: 102582.

Cole, Shawn A., and A. Nilesh Fernando. 2021. "'Mobile'izing Agricultural Advice Technology Adoption Diffusion and Sustainability." *Economic Journal* 131 (633): 192–219.

Corral, Carolina, Xavier Giné, Aprajit Mahajan, and Enrique Seira. 2020. "Autonomy and Specificity in Agricultural Technology Adoption: Evidence from Mexico." NBER Working Paper 27681.

Courtois, Pierre, and Julie Subervie. 2015. "Farmer Bargaining Power and Market Information Services." *American Journal of Agricultural Economics* 97 (3): 953–77.

Cranados, G., S. Pandey, and H. Ceballos. 1993. "Response to Selection for Tolerance to Acid Soils in a Tropical Maize Population." *Crop Science* 33 (5): 936–40.

Cunha, Nina, Guilherme Lichand, Ricardo Madeira, and Eric Bettinger. 2017. "What Is It about Communicating with Parents?" Unpublished.

Dahal, Mahesh, and Nathan Fiala. 2020. "What Do We Know about the Impact of Microfinance? The Problems of Statistical Power and Precision." *World Development* 128: 104773.

Deeks, Jonathan J. 2002. "Issues in the Selection of a Summary Statistic for Meta-analysis of Clinical Trials with Binary Outcomes." *Statistics in Medicine* 21 (11): 1575–1600.

Deichmann, Uwe, Aparajita Goyal, and Deepak Mishra. 2016. "Will Digital Technologies Transform Agriculture in Developing Countries?" *Agricultural Economics* 47 (S1): 21–33.

DellaVigna, Stefano, and Elizabeth Linos. 2022. "RCTs to Scale: Comprehensive Evidence from Two Nudge Units." *Econometrica* 90 (1): 81–116.

DerSimonian, Rebecca, and Nan Laird. 1986. "Meta-analysis in Clinical Trials." *Controlled Clinical Trials* 7 (3): 177–88.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2008. "How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya." *American Economic Review* 98 (2): 482–88.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review* 101 (6): 2350–90.

Dustan, Andrew, Juan Manuel Hernandez-Agramonte, and Stanislao Maldonado. 2023. "Motivating Bureaucrats with Behavioral Insights When State Capacity is Weak: Evidence from Large-Scale Field Experiments in Peru." *Journal of Development Economics* 160: 102995.

Engels, Eric A., Christopher H. Schmid, Norma Terrin, Ingram Olkin, and Joseph Lau. 2000. "Heterogeneity and Statistical Significance in Meta-analysis: An Empirical Study of 125 Meta-analyses." *Statistics in Medicine* 19 (13): 1707–28.

Fabregas, Raissa, Michael Kremer, and Frank Schilbach. 2019. "Realizing the Potential of Digital Development: The Case of Agricultural Advice." *Science* 366 (6471).

Fabregas, Raissa, Michael Kremer, and Frank Schilbach. 2023. "The Effectiveness of Public Agricultural Extension: Evidence from Farmer Field Days in Kenya." Unpublished.

Fabregas, Raissa, Michael Kremer, Matthew Lowes, Robert On, and Giulia Zane. 2020a. *Digital Information Provision and Farmer Behavior: Lessons from Six Experiments in East Africa*. AEA RCT Registry. https://doi.org/10.1257/rct.6837-1.0.

Fabregas, Raissa, Michael Kremer, Jonathan Robinson, and Frank Schilbach. 2020b. "The Value of Local Agricultural Information: Evidence from Kenya." Unpublished.

Fabregas, Raissa, Michael Kremer, Matthew Lowes, Robert On, and Giulia Zane. 2025. *Data and Code for: "Digital Information Provision and Farmer Behavior: Lessons from Six Experiments in East Africa."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E186241V1.

Fafchamps, Marcel, and Bart Minten. 2012. "Impact of SMS-Based Agricultural Information on Indian Farmers." *World Bank Economic Review* 26 (3): 383–414.

FAO. 2015. *Status of the World's Soil Resources (SWSR)–Main Report*. Rome, Italy: Food and Agriculture Organization of the United Nations.

Fleiss, Joseph L., and Jesse A. Berlin. 2009. "Effect Sizes for Dichotomous Data." In *Handbook of Research Synthesis and Meta-analysis*, Vol. 2, edited by Harris Cooper, Larry V. Hedges, and Jeffrey C. Valentine, 237–53. New York, NY: Russel Sage Foundation.

Friedman, Willa, Benjamin Woodman, and Minki 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 (S1): i82–i92.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 1995. *Bayesian Data Analysis*. London, UK: Chapman and Hall/CRC.

Gillwald, Alison, and Onkokame Mothobi. 2019. *After Access 2018: A Demand-Side View of Mobile Internet from 10 African Countries*. Cape Town: Research ICT Africa.

Glewwe, Paul, Michael Kremer, Sylvie Moulin, and Eric Zitzewitz. 2004. "Retrospective versus Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya." *Journal of Development Economics* 74 (1): 251–68.

Global Administrative Areas. 2012. *Global (Database). GADM Database of Global Administrative Areas, Version 2.0 (Online)*. GADM. www.gadm.org (accessed April 15, 2024).

Grossman, Guy, Macartan Humphreys, and Gabriella Sacramone-Lutz. 2020. "Information Technology and Political Engagement: Mixed Evidence from Uganda." *Journal of Politics* 82 (4): 1321–36.

GSMA. 2020. *10 Years of Mobile for Development*. London, UK: GSMA. https://www.gsma.com/mobilefordevelopment/10yearsofm4d/ (accessed January 2021).

Gudu, S. O., J. R. Okalebo, C. O. Othieno, P. A. Obura, D. O. 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*, Vol. 7, 1109–15. Kampala, Uganda: African Crop Science Society.

Gupta, Apoorv, Jacopo Ponticelli, and Andrea Tesei. 2020. "Information, Technology Adoption, and Productivity: The Role of Mobile Phones in Agriculture." NBER Working Paper 27192.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. 2023. "Designing Information Provision Experiments." *Journal of Economic Literature* 61 (1): 3–40.

Hall, Bronwyn H., and Beethika Khan. 2003. "Adoption of New Technology." Unpublished.

Hall, Charles S., Edward Fottrell, Sophia Wilkinson, and Peter 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.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. 2014. "Learning through Noticing: Theory and Evidence from a Field Experiment." *Quarterly Journal of Economics* 129 (3): 1311–53.

Higgins, Julian P. T., and Sally Green. 2008. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester, UK: John Wiley and Sons.

Higgins, Julian P. T., and Simon G. Thompson. 2002. "Quantifying Heterogeneity in a Meta-analysis." *Statistics in Medicine* 21 (11): 1539–58.

Higgins, Julian P. T., Simon G. Thompson, Jonathan J. Deeks, and Douglas G. Altman. 2003. "Measuring Inconsistency in Meta-analyses." *BMJ* 327 (7414): 557–60.

Ioannidis, John P. A., T. D. Stanley, and Hristos Doucouliagos. 2017. "The Power of Bias in Economics Research." *Economic Journal* 127 (605): F236–F265.

Jamison, Julian C., Dean Karlan, and Pia Raffler. 2013. "Mixed-Method Evaluation of a Passive mHealth Sexual Information Texting Service in Uganda." *Information Technologies and International Development* 9 (3): 1–28.

Jensen, Robert. 2007. "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector." *Quarterly Journal of Economics* 122 (3): 879–924.

Kanyanjua, S. M., L. Ireri, Stephen Wambua, and S. M. Nandwa. 2002. "Acidic Soils in Kenya: Constraints and Remedial Options." KARI Technical Note 11.

Karlan, Dean, and Jonathan Zinman. 2008. "Lying about Borrowing." *Journal of the European Economic Association* 6 (2–3): 510–21.

Karlan, Dean S., and Jonathan Zinman. 2012. "List Randomization for Sensitive Behavior: An Application for Measuring Use of Loan Proceeds." *Journal of Development Economics* 98 (1): 71–75.

Karlan, Dean, Melanie Morten, and Jonathan Zinman. 2015. "A Personal Touch in Text Messaging Can Improve Microloan Repayment." *Behavioral Science and Policy* 1 (2): 25–31.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. 2016. "Getting to the Top of Mind: How Reminders Increase Saving." *Management Science* 62 (12): 3393–3411.

Kelly, Valerie, and Anastase 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.

Kisinyo, P. O., P. A. Opala, V. Palapala, S. O. Gudu, C. O. Othieno, and E. Ouma. 2015. "Microdosing of Lime, Phosphorus, and Nitrogen Fertilizers Effect on Maize Performance on an Acid Soil in Kenya." *Sustainable Agriculture Research* 4 (2): 21–30.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119.

KSHC. 2014. "Policy Brief: Soil Health Research Gaps in Kenya." KSHC. https://www.kalro.org/sites/default/files/KSHC (accessed January 2015).

Larochelle, Catherine, Jeffrey Alwang, Elli Travis, Victor Hugo Barrera, and Juan Manuel Dominguez Andrade. 2019. "Did You Really Get the Message? Using Text Reminders to Stimulate Adoption of Agricultural Technologies." *Journal of Development Studies* 55 (4): 548–64.

Lobell, David B., George Azzari, Marshall Burke, Sydney Gourlay, Zhenong Jin, Talip Kilic, and Siobhan Murray. 2020. "Eyes in the Sky, Boots on the Ground: Assessing Satellite-and Ground-Based Approaches to Crop Yield Measurement and Analysis." *American Journal of Agricultural Economics* 102 (1): 202–19.

Marenya, Paswel P., and Christopher B. Barrett. 2009. "State-Conditional Fertilizer Yield Response on Western Kenyan Farms." *American Journal of Agricultural Economics* 91 (4): 991–1006.

McKenzie, David, and Christopher Woodruff. 2014. "What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World?" *World Bank Research Observer* 29 (1): 48–82.

Meager, Rachael. 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, Sandip, Dilip Mookherjee, Maximo Torero, and Sujata Visaria. 2018. "Asymmetric Information and Middleman Margins: An Experiment with Indian Potato Farmers." *Review of Economics and Statistics* 100 (1): 1–13.

Muralidharan, Karthik, Mauricio Romero, and Kaspar Wüthrich. Forthcoming. "Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments." *Review of Economics and Statistics***.**

NAAIAP. 2014. *Soil Suitability Evaluation for Maize Production in Kenya*. Nairobi, Kenya: IFDC Kenya. https://africafertilizer.org/wp-content/uploads/2017/04/Soil-Suitability-Evaluation-for-Maize-Prod.-in-Kenya.pdf (accessed June 2016).

Nakasone, Eduardo, Maximo Torero, and Bart Minten. 2014. "The Power of Information: The ICT Revolution in Agricultural Development." *Annual Review Resource Economics* 6 (1): 533–50.

Oreopoulos, Philip, and Uros Petronijevic. 2019. "The Remarkable Unresponsiveness of College Students to Nudging and What We Can Learn from It." NBER Working Paper 26059.

Pritchett, Lant, and Justin Sandefur. 2015. "Learning from Experiments When Context Matters." *American Economic Review* 105 (5): 471–75.

Raifman, Julia R. G., Heather E. Lanthorn, Slawa Rokicki, and Günther Fink. 2014. "The Impact of Text Message Reminders on Adherence to Antimalarial Treatment in Northern Ghana: A Randomized Trial." *PloS one* 9 (10): e109032.

Rosenzweig, Mark R., and Christopher Udry. 2020. "External Validity in a Stochastic World: Evidence from Low-Income Countries." *Review of Economic Studies* 87 (1): 343–81.

Rubin, Donald B. 1981. "Estimation in Parallel Randomized Experiments." *Journal of Educational Statistics* 6 (4): 377–401.

Schmidt, Susanne, and Martin Eisend. 2015. "Advertising Repetition: A Meta-analysis on Effective Frequency in Advertising." *Journal of Advertising* 44 (4): 415–28.

Sheahan, Megan, and Christopher B. Barrett. 2017. "Ten Striking Facts about Agricultural Input Use in Sub-Saharan Africa." *Food Policy* 67: 12–25.

Steinhardt, Laura C., Don P. Mathanga, Dyson Mwandama, Humphreys Nsona, Dubulao Moyo, Austin Gumbo, Miwako Kobayashi, 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." *American Journal of Tropical Medicine and Hygiene* 100 (2): 460–69.

Suri, Tavneet. 2011. "Selection and Comparative Advantage in Technology Adoption." *Econometrica* 79 (1): 159–209.

Svensson, Jakob, and David Yanagizawa. 2009. "Getting Prices Right: The Impact of the Market Information Service in Uganda." *Journal of the European Economic Association* 7 (2–3): 435–45.

The, Charles, Henri Calba, Célicard Zonkeng, Eddy Léonard M. Ngonkeu, Victor O. Adetimirin, Hortense A. Mafouasson, S. S. Meka, and Walter J. Horst. 2006. "Response of Maize Grain Yield to Changes in Acid Soil Characteristics after Soil Amendment." *Plant and Soil* 284: 45–57.

Tisdale, Samuel L., Werner L. Nelson, and James D. Beaton. 1990. *Soil Fertility and Fertilizers*. New York, NY: Macmillan.

Tjernström, Emilia, Travis J. Lybbert, Rachel Frattarola Hernández, and Juan Sebastian Correa. 2021. "Learning By (Virtually) Doing: Experimentation and Belief Updating in Smallholder Agriculture." *Journal of Economic Behavior and Organization* 189: 28–50.

UNESCO. 2022. *Institute for Statistics* (*UIS*) *Database*. Paris, France: UNESCO. http://data.uis.unesco.org (accessed May 2023).

Van Campenhout, Bjorn, David J. Spielman, and Els Lecoutere. 2021. "Information and Communication Technologies to Provide Agricultural Advice to Smallholder Farmers: Experimental Evidence from Uganda." *American Journal of Agricultural Economics* 103 (1): 317–37.

Vivalt, Eva. 2020. "How Much Can We Generalize from Impact Evaluations?" *Journal of the European Economic Association* 18 (6): 3045–89.