De-Sludging Digital Public Service Delivery: A Data Analytics Approach*

Torsten Figueiredo Walter† Michael Kremer‡ Ofir Reich§ Zhengyun Sun¶
Sam van Herwaarden‖ Habtamu Yesigat** Elias Nure†† Temesgen Gebeyehu‡‡

August 19, 2026

Abstract

This paper proposes a data analytics approach for identifying and reducing frictions in digital public service delivery and evaluates its effectiveness in the context of Ethiopia's 8028 farmer hotline, an interactive voice response (IVR) system for agricultural advice. Guided by funnel analysis identifying IVR menus with high user attrition and by a user survey and focus groups investigating the causes of that attrition, seven system modifications were developed and tested through iterative A/B tests. Three modifications, all of which reduced user hassle, significantly lowered attrition and increased information delivery. The remaining four modifications did not increase information delivery. The most effective intervention addressed friction associated with the collection of user information: removing a multi-step registration process increased access to agricultural advice by 18% without sacrificing valuable user data. In response to these findings, the government ultimately eliminated counterproductive data collection requirements, improving overall system performance. These results demonstrate that data analytics can identify and reduce administrative "sludge" in digital public services and show how data-driven feedback loops can support continuous improvements in government service delivery.

Keywords: A/B Testing, Agricultural Extension, Funnel Analysis, Public Service Delivery, Sludge
JEL Codes: C93, H83, Q16

*This study was funded by Precision Development (PxD) and PxD was directly involved in the study through the affiliation of the majority of the authors with PxD. Michael Kremer is a board member and co-founder of PxD, and Habtamu Yesigat is a senior research and operations manager at PxD. Ofir Reich and Sam van Herwaarden worked as data scientists at PxD when the experiments described in this paper were carried out. Torsten Figueiredo Walter's post-doctoral fellowship was funded by PxD. Elias Nure is the former team leader for ICT for Agriculture at ATA, and Temesgen Gebeyehu is the current Director for Digital Agriculture at ATA. We thank PxD and ATA for their support on executing experiments and collecting data. We thank Daniel Bjorkegren for valuable comments.

†Corresponding author. Division of Social Science, New York University Abu Dhabi. Email: t.f.walter@nyu.edu. Postal address: New York University Abu Dhabi, Building A5, Office 193, P.O. Box 129188, Abu Dhabi, UAE.
‡Department of Economics, University of Chicago and NBER. Email: Kremerassistant@uchicago.edu.
§Independent researcher. Email: ofir.re@gmail.com.
¶Department of Economics, Harvard University. Email: zsun@g.harvard.edu.
‖Precision Agriculture for Development. Email: svanherwaarden@precisionag.org.
**Precision Agriculture for Development. Email: hyesigat@precisionag.org.
††Mercy Corps AgriFin. Email: enure01@gmail.com.
‡‡Ethiopia Agricultural Transformation Agency. Email: Temesgen.Gebeyehu@ata.gov.et.

1 Introduction

Low quality of public services has long been a concern among social scientists and practitioners alike. An extensive literature in economics has shed light on service failures by analyzing rent-seeking motives and accountability mechanisms along the service delivery chain, from the political elite (Acemoglu & Robinson 2012) down to frontline providers (Finan et al. 2015). More recently, however, organizational inertia unrelated to rent-seeking motives of political actors and service delivery agents has been highlighted as a complementary explanation (Bandiera et al. 2009, DellaVigna et al. 2022). Particular emphasis has been put on the seemingly non-strategic nature of many frictions in service delivery systems, also referred to as sludge, with some arguing that service delivery systems tend to suffer from an absence of deliberate design rather than strategically designed obstacles, given the low priority of implementation issues on the political agenda (Bergman et al. 2020, Pahlka 2023, Sunstein 2022).[1]

If many implementation frictions are indeed non-strategic, removing them may be comparatively feasible because doing so does not threaten entrenched interests. At the same time, their persistence suggests that governments may simply not know where they occur or how consequential they are. The challenge, therefore, is to identify where these frictions arise, understand why they impede users, and determine which potential remedies are effective. The growing digitization of public service delivery creates new opportunities to do so. Digital systems generate granular data on user behavior and facilitate rapid testing of service modifications, creating opportunities to apply data analytics tools commonly used in the technology sector to identify frictions and evaluate potential improvements.

This paper develops and applies a data analytics approach for identifying and reducing service delivery frictions in an interactive voice response (IVR) system for agricultural advice. The approach uses quantitative funnel analysis of system data, a user survey, and focus groups to identify points of high user attrition and investigates their underlying causes, and then develops and iteratively A/B tests potential system improvements. Of seven system modifications tested, three lowered user attrition and increased information delivery. The most effective intervention removed a multi-step registration process, increasing access to agricultural advice by 18%. These findings ultimately prompted the government to remove counterproductive data collection requirements, indicating that data-driven feedback loops can support continuous improvements in government service delivery.

The phone-based agricultural extension system at the center of this paper is Ethiopia's 8028 farmer hotline, a flagship government program created by the Agricultural Transformation Agency (ATA), the Ministry of Agriculture & Livestock Resources, the Ethiopian Institute of Agricultural Research and Ethio Telecom in 2014. It is an IVR system that allows farmers to access pre-recorded

[1] This is not to say that there are no cases where sludge is imposed strategically. In fact, there are well known examples of strategic sludge (Moynihan & Herd 2019, Chapter 2). However, a substantial share of sludge does not appear to serve the strategic interest of any particular stakeholder.

agricultural advice messages through menu selection (e.g., "For information on post-harvest and processing, press 5."). To date, it has recorded 7.1 million users (Capital Ethiopia 2026).

Though the hotline is broadly regarded as a success and a substantial share of farmers seek advice through the system, many users hang up before they can access advice in the IVR menu, with only 37% of first-time callers succeeding in accessing any advice. We applied funnel analysis to call logs to pinpoint areas of severe user attrition within the IVR menu, conducted focus groups and a user survey to understand the associated barriers, and developed system modifications to address these. The impact of seven system changes thus designed was evaluated through 10 A/B tests — randomized control trials that compare user attrition between those exposed to the modified system versus the baseline system.

The most effective interventions reduced frictions due to the collection of user data. In particular, eliminating a lengthy user registration process increased agricultural advice accessed by 10-18%. At the same time, the cost of foregoing user registration data collection was low as information provided by users was largely inaccurate and failed to serve the purpose of targeted messaging based on farmer characteristics. A second successful intervention removed a user prompt to save a selected crop to their profile, increasing content accessed by up to 5.8% while crop choice remains observable through call logs. Of the other five interventions that were tested, one increased information access by 2.5% and the other four did not prove effective, highlighting the importance of A/B testing for scrutinizing the performance of a supposed improvement in practice.

Two common threads link these findings. First, interventions that reduced frictions along the menu path led to increases in content accessed while interventions that introduced additional steps along the menu path, although meant to improve user experience, did not have a positive impact. Second, frictions generated by the collection of user information were particularly detrimental to content access. Removing these implied the largest gains, and any attempts to partially re-introduce user registration failed.

The government asked for multiple trials and for an exploration of alternatives to reducing data collection, but eventually decided to reduce information requests to users in light of the evidence. The ultimate decision to eliminate counterproductive data collection suggests that the existence of this sludge was not strategic, i.e., not serving as a deliberate ordeal designed to exclude certain populations from public services (Moynihan & Herd 2019) or as a part of a deliberate rent-seeking effort by politicians or corrupt public servants seeking to advance their personal financial interests. Instead, it seems more related to the type of "passive" government failure recently emphasized in the literature (Bandiera et al. 2009, DellaVigna et al. 2022, Pahlka 2023). It also suggests that governments may be responsive to evidence, consistent with Hjort et al. (2021) and Callen et al. (2020). It thus points to a potential role for this type of data analytics approach to identify and reduce sludge in digital public service delivery systems by using qualitative data collection, funnel analysis, and A/B testing to generate and evaluate ideas for system improvements, just as similar approaches are used in the tech sector (Tang et al. 2010, Dumais

et al. 2014, Kohavi et al. 2020, Koning et al. 2022).

That said, the government's requests for multiple trials and for an exploration of alternatives prior to reducing data collection suggest that data analytic efforts may only affect policy if they are iterative and ongoing. The analytics were externally supported, and while the government ultimately adopted the approaches found to be most successful in the A/B tests, it did not go on to use its own resources to set up a permanent unit to analyze system data and test potential improvements, including potential improvements based on reducing data collection. Our analysis is thus also consistent with Sunstein's (2022) view that individual governments may need to create overarching regulations and laws and government units to limit sludge, such as the Paperwork Reduction Act and the Office of Information and Regulatory Affairs (OIRA) in the US.

This paper contributes to four streams of literature. First, it contributes to a large literature that emphasizes the adverse effects of sludge in public service delivery systems (Sunstein 2022), particularly its role in explaining low take-up of public benefit programs and foregone tax savings (Bettinger et al. 2012, Bhargava & Manoli 2015, Benzarti 2020). It extends this by demonstrating the importance of sludge in the realm of digital public service delivery and by proposing a systematic data-driven approach to reducing it that integrates empirical tools from the social sciences and computer science.

Second, this paper adds to a growing set of case studies that support the idea of data-driven feedback loops for government services as a complement to feedback loops from citizens through political action or client power, which may be hampered by low levels of accountability (Pritchett et al. 2013, Gelb et al. 2019). While most existing work focuses primarily on using administrative data and user surveys to monitor service delivery (Muralidharan & Singh 2020, Ganimian et al. 2024), this paper emphasizes how data analysis, focus groups and A/B testing can be used to inform the design of public service delivery.

Third, this paper speaks to a growing literature on digital service delivery. While evidence of the cost-effectiveness of digital services for agriculture and education has been mounting in recent years (Casaburi et al. 2019, Fabregas et al. 2019, Cole & Fernando 2021, Angrist et al. 2022, Fabregas et al. 2025, Angrist et al. 2025), low take-up and high attrition remain a concern (Fafchamps & Minten 2012, Aker et al. 2016, Van Campenhout et al. 2021, Figueiredo Walter et al. 2023). This paper adds to this literature by developing and testing a systematic approach to identifying and overcoming usability barriers that limit retention. Additionally, it connects the study of digital advisory services to the well-established concept of funnel optimization in the tech sector, which emphasizes minimizing the time and effort required for users to reach the "magic moment" of successfully accessing a service (Ellis & Brown 2017).

Finally, this paper relates to a nascent literature on adaptive experimentation as a tool to enhance digital services (Kasy & Sautmann 2021, Athey 2025, Viviano & Rudder 2026). Similar to Angrist et al. (2025) and Athey et al. (2026), it demonstrates how iterative A/B testing can be leveraged to improve digital service delivery.

The remainder of the paper is structured as follows. Section 2 provides background informa-

tion on Ethiopia's IVR system for agricultural advice. Section 3 presents funnel analysis results from user call logs and descriptive evidence from a complementary telephone survey. Section 4 documents the iterative development and testing of system modifications. Section 5 discusses potential mechanisms underlying the core results, and Section 6 concludes.

2 Background

The Ethiopian Agricultural Transformation Agency (ATA) was established in 2010. ATA is a public agency with the aim to support transformation of smallholder agriculture into a productive and fast-growing sector. As such, it conducts studies to identify systemic bottlenecks and priorities in the sector and recommends solutions. In addition, it implements pilot interventions that aim to support the transformation process. Among these interventions is the 8028 farmers' hotline system for agricultural advice.

The 8028 farmers' hotline system was established to provide a timely and reliable extension service at scale, complementing Ethiopia's network of approximately 60,000 development agents providing in-person extension services. At the core of the hotline is an interactive voice response system that allows farmers to access pre-recorded agricultural advice messages. Consistent with some views of bureaucratic processes in governments, the program is fairly comprehensive and thus complex, rather than just focusing on a few use cases. The content available on the IVR platform covers 19 grain and vegetable crops in 5 major languages. As of today, the 8028 farmers' hotline system is one of the largest digital agricultural extension systems in Africa, with more than 7.1 million callers since its launch and the potential to reach millions more.

From July 2017 on, ATA and Precision Development (PxD) collaborated to improve the 8028 farmers' hotline system. Joint activities have included system diagnosis to identify potential for performance gains, design and testing of potential improvements and the permanent implementation of successfully tested system changes. This paper reports on some of this joint work.

Hotline users must provide information in an extensive series of IVR steps before they can access any content. First-time users begin by selecting their language and creating a user profile. To complete profile registration, users must provide their caller type specification (farmer, model farmer, extension worker, etc.), location (region, zone, district), gender and land size over two calls.[2] Registration is followed by the *Top Menu* where users are required to provide information on whether they want content on rain-fed and irrigated crops and are subsequently taken to the corresponding sub-menus, in which they supply information on the activity (e.g., planting, crop protection, harvesting) and the crop they are interested in.[3] Finally, before listening to content, they are asked to save their crop choice to their profile. After the content has been played, they can either choose to replay the message they have listened to, or select another crop, or return to

[2]The registration process is conducted in two steps. In their first call, users are asked to specify their caller type, region and zone. In their second call, their district, gender and land size are elicited. Detailed flowcharts of the overall call process and the registration menu in particular are provided in Figures A.1, B.1 and B.2.

[3]See Figures B.3 and B.4 for detailed flowcharts of these sub-menus.

the top menu. At any point they can also terminate the call by hanging up.

The content is developed and verified by agronomists. For example, the IVR platform recommends farming techniques and seed varieties for wheat that have been found to significantly increase yields in a randomized control trial (Abate et al. 2018). Content is not customized to farmers based on their characteristics (e.g., location). Instead, messages typically cover multiple scenarios to accommodate the large variation in agro-climatic conditions across the country. The language tends to be rather technical. Three exemplary messages are provided in Appendix A.1.

3 Funnel analysis and telephone survey

To understand the experience of users with the IVR system, we analyzed call logs which record every navigation step as a separate event.[4] For this purpose, we drew a random sample of 50,000 users from all users that had called into the system by the end of July 2017 and extract their entire call logs up to that point. Then we focused on all users that made their first call between January 2016 and May 2017.[5] The final sample for analysis contained 17,378 users who made 123,123 calls.

Our analysis revealed that many users did not advance very far in the IVR menu. As figure 1 shows, in their first call 16% of users did not even select a language, only 48% completed the first part of registration (i.e., reach the Top Menu) and a mere 37% listened to any content. Among the users who did not complete the first part registration on their first call, 64% called again and completed this part within the next two months. But the remaining users — 18% of all users — never completed it within the first two months following their first call and did not access any content in this time period. Thus, a substantial share of users failed to obtain any agricultural advice from the system.

Having completed the first part of registration, 68% of users called again. At the mean, these users made seven additional calls over the following two months and listened to equally many content items. However, the distributions of additional calls and content items were very skewed.[6] At the median, these users only made four additional calls and listened to three content items. More than 25% of them made no more than two additional calls and only 82% completed the second part of the registration process within the following two months. Hence, even among those who had completed the first part of the registration process, many were not retained.

Why was retention so low? Was there simply no demand for further information? Or did users face barriers interacting with the system? To shed light on these questions, we conducted a complementary telephone survey on farming practices and IVR system usage among a random sample of 1,403 users that were registered as (model) farmers and had called at least once in

[4] See Figure A.2 for an illustration.

[5] We selected this time interval because the IVR operated unchanged over this period. Before 2016 and after May 2017, the user registration process was slightly different from the one described in the previous section.

[6] See Figure A.3 for the distributions of calls and content items listened to in the first two months following the completion of the first part of registration.

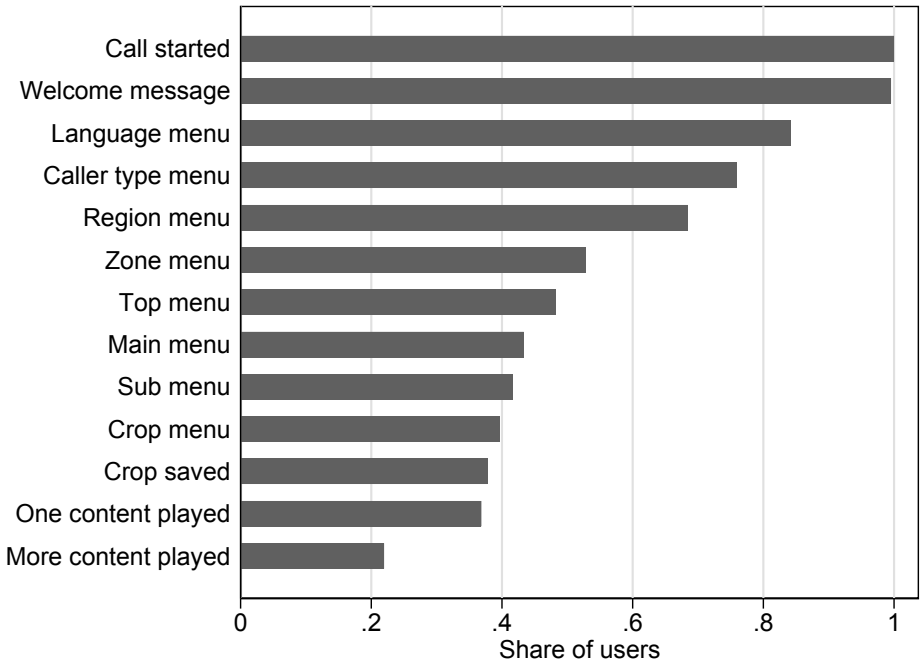


Figure 1: User progress through consecutive menus on first call

This figure shows the share of users reaching each consecutive menu step on their first call. It is constructed from a random sample of 17,378 callers who made their first call between January 2016 and May 2017.

the last five months.[7] When asked about barriers to using the IVR system effectively, farmers reported substantial difficulties interacting with the system: 74% of the survey respondents reported problems operating the system (Figure A.4). These difficulties also became apparent when the location information entered into the IVR system was compared with survey responses. The share of respondents providing consistent region information in the system and the survey was only 75%. The corresponding figures for zone and district information were even smaller, 48% and 18%, respectively. Thus, the available evidence pointed to user difficulties with the IVR system rather than lack of demand for advice as the underlying cause for low content access and user retention.[8]

On the bright side, this suggested that improved system design could lead to increased content access and retention. In the remainder of the paper, we show explicitly that there were indeed simple system changes that significantly increased content access in the short and medium run.

Note that we did not address barriers to retention resulting from the design of the content messages themselves in this paper, even though these might also have been important. After all,

[7] See Online Appendix B.3 for details on the survey.

[8] An alternative interpretation of the inconsistency in location information provided by users was that users intentionally entered incorrect information due to mistrust towards the government. However, when asked in the survey, most users (94%) indicated that they trust the recommendations given by the IVR system. Moreover, the degree of trust was positively correlated with inconsistency in location information across survey respondents - not negatively, as one might expect in the case of intentionally misspecified information due to a lack of trust.

about one fourth of survey respondents indicated difficulties understanding and implementing the rather technical recommendations (Figure A.4).

4 A/B testing

4.1 Research design

4.1.1 Iterative testing

Ten A/B tests were carried out sequentially and iteratively across four testing rounds. First-round A/B tests were based on the analysis of historical system data, while later experiments either re-tested previous interventions, tested modified versions of earlier interventions, or examined the effectiveness of new interventions whose design was informed by previous findings. Experiments focused on profile registration and menu navigation. Table A.1 provides an overview of all tests conducted across the four rounds.

4.1.2 Outcome measures

Our primary outcome of interest is an indicator variable that takes the value one if a user accessed at least one content message in the first two months following their initial call, and zero otherwise.[9] We also evaluate the impact of all interventions on two other measures of content accessed by users: the total number of content items accessed and the number of unique content items accessed. While the former counts an identical message that a user listens to multiple times, the latter counts it only once. Additionally, for individual experiments we assess the effect on auxiliary outcomes that shed light on underlying mechanisms. For each experiment, these are described in more detail below.

Table A.2 presents summary statistics for all variables across all seven data samples used for the evaluation. For each experiment, we use the relevant experimental sample for evaluation, as indicated in the second row of the table.

4.1.3 Empirical strategy

Throughout, we estimate the average treatment effect on accessing content as well as auxiliary outcomes using the following linear probability model, separately for each experiment:

$$y_i = \alpha + \beta T_i + \epsilon_i \tag{1}$$

[9]The central nature of this outcome was specified prior to the first round of experiments and its selection was based on three arguments: first, its primary importance and immediate link to the analytical results presented in section 3 (after all, listening to content is the prerequisite for obtaining any information from the platform); second, its simplicity; third, its benefits in terms of power relative to less discrete outcome measures (Kohavi et al. 2009). However, none of the experiments reported on in this paper was formally pre-registered as the scope of our work with ATA was initially unclear and only became apparent as experimentation progressed.

where T_i is an indicator that takes value one if user i was treated in the experiment, and zero otherwise. y_i is a measure of the content accessed by user i or an auxiliary outcome variable.

The results from the reported short regressions formed the basis for government decisions about the abortion, adoption and further testing of interventions. Following Muralidharan et al. (2019), we also provide the results from long regressions including indicator variables for all treatment combinations in Online Appendix B.1 and show that these results equally support all the decisions taken by ATA. Evidence supporting experimental integrity is provided in the form of balance tests in Table A.3 and further discussed in Appendix A.2.

4.2 Preview of results

The flowchart in Figure 2 visualizes the iterative A/B testing process and its results. Experiments evolved around the five topics indicated at the top of the chart: user registration, saving the crops selected by users, menu replay, pauses between menu options and a user introduction to the IVR platform. In total, ten experiments were conducted over four rounds. Interventions are described in purple boxes, effects on content accessed and intermediate outcomes of interest are summarized in white boxes and adoption decisions are highlighted in green (adopt), yellow (further experimentation) and red (discard) boxes. Details on all the experiments, results and adoption decisions are provided in the following sections.

Without going into detail on any of the experiments, one feature stands out. Even if the experimental results were positive, system changes were rarely adopted immediately. Instead, successful tests were often followed by further tests of alternative system changes. This is particularly true in the case of user registration, reflecting hesitancy on the side of the government agency to forgo the collection of user data. Adoption typically only occurred after repeatedly ensuring the positive impact of an intervention. Negative test results, on the other hand, led to immediate cancellation of testing and discarding of interventions (e.g., intro to IVR).

4.3 First round of experiments

In the first round of experiments, a total of four experiments were conducted. Details on all four experiments are provided below. Results are summarized in Table 1.[10]

Experiment 1.1: Removing registration. Agricultural advice provided by the hotline is not customized to users based on information provided in the registration process. However, user profiles are stored with the aim to enable the platform provider to send targeted messages to farmers, e.g., location-specific alerts on weather or pests. But the benefits of acquiring this information are likely to be limited because the accuracy of the information provided in the registration process is low,

[10] Appendix tables B.1 and B.2 show the results from the corresponding long regressions for experiments 1.1-1.3 and experiment 1.4, respectively. Unless mentioned otherwise, the results reported in the appendix are qualitatively and quantitatively similar to the ones shown in the main table.

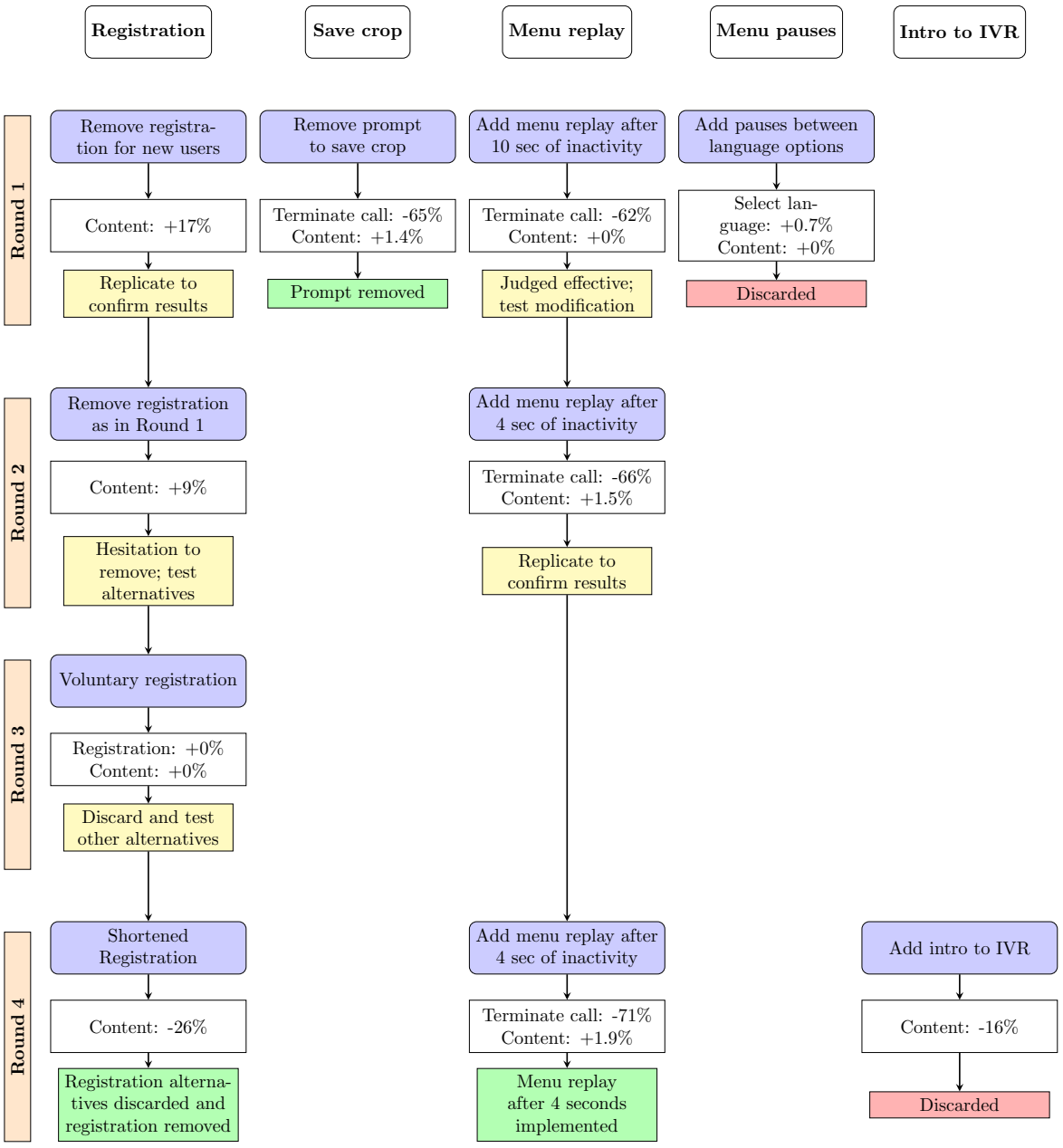


Figure 2: Flowchart of iterative A/B testing process

This figure illustrates the iterative A/B testing process across experimentation rounds (left) and domains (top). Titles of experiments are shown in purple boxes, estimated effects in white boxes and government decisions in red (discarded), yellow (further testing) and green (system-wide adoption) boxes.

as revealed by our telephone survey. Moreover, the registration process itself may pose a major barrier to many users seeking to access content. As shown in section 3, many new users do not progress beyond the initial stages of the registration process and never access any content.

In this experiment, we tested whether registration itself is indeed an important barrier to accessing content. New users were randomly assigned to a treatment and a control group (with equal probability). After selecting a language, new users in the treatment group were not prompted to register but immediately directed to the *Top Menu* from where they could access content.

We find that removing registration has a large positive effect on content accessed. The probability of listening to at least one content message in the first two months increases by 10.2 percentage points (pp) relative to a control mean of 58.6 pp, corresponding to a 17% increase (Panel A, Column 1). Similarly, the number of (unique) content items listened to in the first two months increases by 16% (18%).[11]

Despite the large positive effect of the removal of the registration on content accessed, there was hesitation to remove the registration due to the perceived loss of user information required for the dissemination of location-specific advice, notwithstanding its poor accuracy demonstrated by the phone survey. Thus, it was decided to replicate the experiment in the next round.

Experiment 1.2: Adding menu replay after 10 seconds of inactivity. After 10 seconds of inactivity (no audio being played by the system and no key being pressed by the user), the system automatically hangs up. In the pre-experimental analysis, we had observed that 4% of first calls were terminated by the system without users completing language selection. Similarly, 6% of calls that reached the top menu were terminated by the system without a choice being made in this menu. One possible explanation is that system options are read out too fast, and users forget the options. In fact, in our telephone survey, 4% of respondents listed quickly spoken menus as one of the three main issues they face in using the IVR system.

This experiment tested whether replaying the menu if no selection is made decreases the termination of calls by the system. New users were randomly assigned to a treatment and a control group (with equal probability). Unlike in the control group, users in the treatment group were not hung up on by the system if they were inactive for 10 seconds after the language or top menu had been played. Instead, these menus were replayed up to twice more before the system hung up automatically.

We find that the treatment leads to a substantial reduction in users having their calls terminated by the IVR system when they reach the language menu or the top menu. In the control group, 7.1% of users are hung up on by the system when reaching the language menu at least once in their first two months. In the treatment group, the share of users that experiences this is

[11] Note that in the first call alone, this effect is even more pronounced, with the removal of the registration causing a 41% increase in content access. In other words, the removal of the registration led to substantially faster content access. Figure A.6 illustrates this graphically. Across call attempts, the share of users that has not yet accessed any content drops much faster among treated than control users. The gap narrows, however, as the number of call attempts increases.

only 1.6% (Panel B, Column 4). The corresponding call termination figures for the top menu are 13.3% in the control group versus 7.3% in the treatment group (Panel B, Column 5). Thus, the relative reductions in call termination by the system are large in both cases, amounting to 78% and 45%, respectively. We barely find any effects on content accessed, however (Panel B, Columns 1-3), suggesting that users face other constraints further downstream in the IVR menu that are binding.

The addition of menu replay was judged as a promising system modification although the experiment did not provide strong evidence of a positive effect of adding menu replay on content accessed. But since a dramatic fall in call termination due to inactivity was observed, the decision was made to run a modified version of this intervention where the time of inactivity before menu replay would be shorter.

Experiment 1.3: Adding pauses between language options. In this experiment, we test another approach to improve the comprehension of menus. We examine how adding pauses of two seconds between options in the language menu affects language selection and content subsequently accessed. New users are randomly assigned to a treatment and a control group (with equal probability). In the treatment group, the above-mentioned pauses between options are introduced while the menu is played without these pauses in the control group.

We do not find any effects on content accessed and even a small negative effect on language selection ($< 1\%$) as shown in Panel C of Table 1 (Column 6). Hence, the introduction of pauses between menu options does not appear to address the problem of quickly spoken menus as voiced by farmers. One potential explanation is that pauses of two seconds come across as artificially long while leaving the speed of speech itself unchanged.

Based on the lack of effect on content accessed and the negative impact on language selection observed in the experiment, the attempt to introduce longer pauses between language options was aborted.

Experiment 1.4: Removing prompt to save crop to profile. The system prompts users to save the selected crop to their profile before listening to content, if the crop has not been added before. The saved information is then used by the platform provider to send crop-specific messages to farmers (e.g., alerts regarding pest outbreaks). However, information on selected crops is also easily available from call logs, thus rendering the crop saving prompt unnecessary for the purpose of message targeting. Moreover, users had reported that the prompt to save their crop was confusing and they did not understand how to respond to the menu options. In line with this, pre-experimental analysis had revealed that 2.5% of the calls in which farmers succeed in selecting a crop are terminated when farmers are prompted to save the crop to their profile subsequently — despite the fact that the system will even move on to play the selected content if users simply hold the line for ten seconds without pressing any key.

This experiment tested whether removing the prompt to save the selected crop to the profile

affects content accessed. Users were randomly assigned to a treatment and a control group (with equal probability). For users in the treatment group, the prompt to save the selected crop to their profile was removed.

We find that the share of users with calls terminated prior to any content having been played, conditional on completed crop selection, is substantially lower in the treatment group. While 7% of users in the control group have calls terminated between crop selection and content replay, only 2.5% of users in the treatment group have calls terminated at this stage, representing a 65% decrease (Panel D, column 7). As a result, treated callers access more content. The probability of accessing any content is 0.9 pp higher in the treatment group, and the number of content items accessed is 0.26 larger, corresponding to a 1.4% and a 5.1% increase relative to the mean in the control group, respectively (Panel D, Columns 1 and 2). Similarly, the number of unique content items accessed is 5.8% higher in the treatment group (Panel D, Column 3). Hence, removing the prompt to save the crop to the profile leads to a small but significant increase in content accessed. Note, however, that this increase is not confirmed by the results from the long regression reported in Table B.2. Relative to pure control users that do not receive any of the four treatments tested in the first round of experiments, users that are only treated in experiment 1.4 (and none of the other three experiments) do not access a significantly larger amount of content. This means that removing the prompt is only effective in the presence of the other treatments, thus pointing to complementarities between different friction reductions along the IVR menu.

Since the removal did not imply any loss of information and no potential downsides to user access to content were detected in the A/B test, it was decided that the prompt to save a crop would be removed.

4.4 Second round of experiments

The second round of experiments included two experiments. All details are provided below and results are shown in Table 2.[12]

Experiment 2.1: Removing registration. This experiment was a replication of experiment 1.1, yielding similar results. Although we find a smaller effect of the removal of registration than in the first trial, effect sizes remain large and significant. We estimate that removing registration leads to a 9% increase in the probability of accessing at least one content item in the first two months after a user's initial call (Panel A, Column 1) and increases in the number of content items listened to are of similar magnitude — 8% (10%) for the number of (unique) content items accessed (Panel A, Columns 2 and 3).

Despite the repeatedly positive test results, it was decided not to abolish registration, but to conduct further testing. In particular, a search for alternative registration processes that would

[12] Appendix table B.3 shows the results from the corresponding long regression, which confirm the findings presented here.

Table 1: Round 1 - Treatment Effects

	Content accessed			System terminated call at ...		Selected ...	Call ended after ...
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Any	# Items	# Unique Items	Lang menu	Top menu	Language	Crop selection
Panel A: Remove registration [Exp. 1.1, Sample 1A]							
Treated	0.102***	0.778***	0.482***				
	(0.004)	(0.083)	(0.038)				
R2	0.011	0.001	0.002				
N	67316	67316	67316				
Control mean	0.586	4.815	2.711				
Panel B: Add menu replay after 10 seconds of inactivity [Exp. 1.2, Sample 1A]							
Treated	0.007*	-0.000	0.008	-0.055***	-0.060***		
	(0.004)	(0.083)	(0.038)	(0.002)	(0.002)		
R2	0.000	0.000	0.000	0.018	0.010		
N	67316	67316	67316	67316	67316		
Control mean	0.634	5.205	2.948	0.071	0.133		
Panel C: Add pauses between language options [Exp. 1.3, Sample 1A]							
Treated	-0.002	-0.080	-0.055			-0.006**	
	(0.004)	(0.083)	(0.038)			(0.002)	
R2	0.000	0.000	0.000			0.000	
N	67316	67316	67316			67316	
Controlm mean	0.639	5.244	2.979			0.891	
Panel D: Remove prompt to save crop [Exp. 1.4, Sample 1B]							
Treated	0.009**	0.261***	0.167***				-0.045***
	(0.004)	(0.083)	(0.038)				(0.002)
R2	0.000	0.000	0.000				0.011
N	68377	68377	68377				68377
Control mean	0.633	5.069	2.865				0.070

See Table A.2 notes for variable definitions. Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

impact user access to content less negatively was initiated.

Experiment 2.2: Adding menu replay after 4 seconds of inactivity. This experiment was a replication of experiment 1.2 with one modification: the time of inactivity before menu replay in the treatment group was reduced from 10 seconds to 4 seconds.

We find a similar reduction in call termination by the system as in the original experiment, amounting to 76% in the language menu and 56% in the top menu (Panel B, Columns 4 and 5). Additionally, we detect a small but statistically significant, positive effect on content accessed. In the treatment group, the share of users accessing any content is 1.5% higher than in the control group and the number of unique content items listened to is 2.5% larger (Panel B, Columns 1 and 3). Thus, the modified treatment not only reduces call termination by the system in the respective menus, but its effect also carries through to the quantity of content accessed.

Nonetheless, the tested system modification was not implemented. Instead, a further verification of the results was desired.

Table 2: Round 2 - Treatment Effects

	Content accessed			System terminated call at ...	
	(1)	(2)	(3)	(4)	(5)
	Any	# Items	# Unique Items	Lang menu	Top menu
Panel A: Remove registration [Exp. 2.1, Sample 2]					
Treated	0.052***	0.376***	0.266***		
	(0.004)	(0.071)	(0.036)		
R2	0.003	0.000	0.001		
N	77628	77628	77628		
Control mean	0.578	4.537	2.712		
Panel B: Add menu replay after 4 seconds of inactivity [Exp. 2.2, Sample 2]					
Treated	0.009***	0.114	0.070**	-0.069***	-0.083***
	(0.004)	(0.071)	(0.036)	(0.002)	(0.002)
R2	0.000	0.000	0.000	0.022	0.018
N	77628	77628	77628	77628	77628
Control mean	0.600	4.667	2.810	0.091	0.148

See Table A.2 notes for variable definitions. Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.5 Third round of experiments

The third round of experiments included only one experiment. Its results are provided in Table 3.

Experiment 3.1: Voluntary registration. Experiments 1.1 and 2.1 had demonstrated the advantages of removing registration from the first call. However, there are also benefits to obtaining profile information, e.g., for sending out targeted pest alerts. This experiment tested whether users are willing to provide profile information voluntarily in a later call.

A random 5% subset of new Amharic-speaking users was randomly assigned to two treatment arms and a control group (with equal probabilities). Users in the first treatment arm were asked profile questions once they had completed a total of five calls, users in the second treatment arm once they had listened to at least three content items. Profile questions were split into two groups of three questions: location (region, zone, district) and other (gender, caller type, land size). On the call following the one in which the minimum requirement had been met, users were asked whether they were willing to answer a set of profile questions. If they agreed to do so, they were asked one group of questions. If not, the call continued as usual. The offer to answer profile questions was repeated every fifth call until all six profile questions had been completed. Users in the control group were not asked any registration questions at any point in time.

This experiment was limited to new Amharic-speaking users to save translation costs for the newly developed registration process. Additionally, the sample was kept small because there was substantial uncertainty about the effects of this intervention and large negative effects at the aggregate level were to be avoided. The trial was designed as an initial test that could be replicated with a larger sample of users from different languages at a later stage if results were promising.

We find no statistically significant effect of treatment on profile registration and content accessed relative to the control group, which was not asked any registration questions. While insignificant, all point estimates are negative. Thus, this intervention appears to have failed to collect additional user data through voluntary registration and may even have reduced the content accessed by callers.[13]

Based on the cumulative evidence on the large positive effect of the removal of registration on content accessed from experiments 1.1 and 2.1, and the failure to collect additional user information through voluntary registration, registration was removed from the IVR menu following this round of experiments. At the same time, further experimentation with alternative registration processes that allow for the collection of basic user data remained a priority.

[13] Results by treatment arm are shown in Appendix table A.5. Effects are neither statistically significantly different from zero nor from each other.

Table 3: Round 3 - Treatment Effects

	Content accessed			Completed ...
	(1) Any	(2) # Items	(3) # Unique Items	(4) Registration
Panel A: Voluntary registration [Exp. 3.1, Sample 3]				
Treated	-0.019	-0.530	-0.352	-0.006
	(0.024)	(0.629)	(0.352)	(0.014)
R2	0.001	0.001	0.001	0.000
N	1072	1072	1072	1072
Control mean	0.839	6.493	4.174	0.057

See Table A.2 notes for variable definitions. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.6 Fourth round of experiments

In the fourth round of experiments, three experiments were conducted. Details on all of these experiments are provided below and results are summarized in Table 4.[14]

Experiment 4.1: Shortened registration. This experiment tested whether a short registration process at the beginning of the first call allows for collection of basic user information without affecting content accessed significantly negatively. New users were randomly assigned to a treatment and a control group (with equal probability). In the treatment group, users underwent a short registration process after having selected their language. They were asked to provide their region, zone and district. Users in the control group were not asked any registration questions.

We find that the shortened registration affects content accessed even more negatively than the original longer registration (relative to the case of no registration). The share of users accessing any content over their first two months as system users is 17pp lower in the treatment group than in the control group, corresponding to a 26% decrease (Panel A, Column 1). Additionally, the number of (unique) content items accessed over the same period is 20% lower (Panel A, Columns 2 and 3). These losses in content accessed are larger than the original gains from removing registration which amounted to 16% to 18%. Thus, the shortened registration process does not resolve the problem of high user attrition through profile registration; it actually aggravates it in our context. One potential explanation for this seemingly counter-intuitive result is that the shortened registration process requires users to choose their district of residence from a long list of menu options on the first call, while this potentially challenging entry is only required on the second call in the long registration process, thus allowing for access to content on the first call

[14]The results from the corresponding long regressions can be found in Appendix tables B.4 and B.5. Unless mentioned otherwise, they are in line with the results from the short regressions.

without the barrier of district entry.[15]

After the fourth round of experiments, the attempt to introduce a shortened registration process was aborted and registration was abolished.

Experiment 4.2: Adding menu replay after 4 seconds of inactivity. This experiment was a replication of experiment 2.2. As in previous versions of this trial, we find evidence of a substantial drop in automatic call termination upon reaching the language or top menu (Panel B, Columns 4 and 5), but only weak evidence of an effect on content accessed (Panel B, Columns 1-3).

Given the repeated evidence in favor of this system modification, the menu replay after 4 seconds of inactivity was finally implemented.

Experiment 4.3: Adding introduction to IVR. Pre-experimental analysis of call logs, focus group discussions and in-depth individual interviews conducted thereafter had revealed that many users were unable to effectively use the system. This experiment tested whether providing a brief introduction to the system and its functionality could improve users' experience and their access to content.

New users who selected the Amharic language were randomly assigned to a treatment and a control group (with probabilities 0.1 and 0.9). Treated users were further assigned to one of two treatment arms (with equal probability). In both treatment arms, on their first call new users were given a short, recorded introduction explaining how to use the IVR system and what its purpose was. The complete introduction took 3 minutes and 17 seconds, but users were allowed to skip the tutorial by pressing "9". The two treatment arms differed with regards to what happened if users pressed some other key. In arm 1, the introduction was skipped. In arm 2, the system responded that the key was invalid and replayed the introduction.[16]

This experiment was limited to new Amharic-speaking users in order to save translation and recording costs for the introduction messages. The plan was to do further testing for other languages if this initial trial were successful. The test was deliberately underpowered due to ex ante concerns about potential adverse effects: treatment assignment was limited to 10% to mitigate the risk of alienating users, at the cost of reduced precision in estimating the treatment effect.

We find a significant negative impact of treatment on content accessed (Panel C, Columns 1-3). As shown in Appendix table A.6, this effect is present for both treatment arms. Across all calls in the first two months, 16% fewer users access any content in the treatment group (Panel C, Column 1). Moreover, the number of (unique) content items accessed is 9% (10%) lower — even if only marginally statistically significantly so (Panel C, Columns 2 and 3). Hence, adding an introduction to the IVR system does not appear to address the concern of limited system understanding successfully. It rather leads to a substantial decrease in content accessed,

[15] We further elaborate on this interpretation in Section 5 when discussing the mechanisms underlying our results across all experiments.

[16] An English translation of the introduction script is available in Appendix A.3.

potentially because it prolongs time to first content access excessively. After all, 3 minutes and 17 seconds is a long time to listen before starting menu navigation.

Based on these results, the idea to play an introductory explanatory message at the beginning of the first call was discarded.

Table 4: Round 4 - Treatment Effects

	Content accessed			System terminated call at ...	
	(1) Any	(2) # Items	(3) # Unique Items	(4) Lang menu	(5) Top menu
Panel A: Short registration [Exp. 4.1, Sample 4A]					
Treated	-0.172***	-0.937***	-0.602***		
	(0.006)	(0.108)	(0.057)		
R2	0.030	0.002	0.004		
N	30517	30517	30517		
Control mean	0.660	4.673	3.023		
Panel B: Add menu replay after 4 seconds of inactivity [Exp. 4.2, Sample 4A]					
Treated	0.011*	-0.031	0.024	-0.058***	-0.092***
	(0.006)	(0.108)	(0.057)	(0.002)	(0.003)
R2	0.000	0.000	0.000	0.020	0.022
N	30517	30517	30517	30517	30517
Control mean	0.568	4.217	2.709	0.072	0.150
Panel C: Add IVR introduction [Exp. 4.3, Sample 4B]					
Treated	-0.115***	-0.460	-0.323*		
	(0.0144)	(0.322)	(0.167)		
R2	0.00554	0.000179	0.000328		
N	11370	11370	11370		
Control mean	0.707	4.973	3.270		

See Table A.2 notes for variable definitions. Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5 Mechanisms

What are the broader lessons we can draw from the above collection of experimental results? In order to understand the mechanisms driving the success of some interventions and the failure of

others, we examine user attrition along IVR menu steps within users' first call. Specifically, we focus on four experiments, two of which succeeded and two of which turned out to be counterproductive: 1.1 (removal of registration), 2.2 (menu replay after 4 seconds), 4.1 (shortened registration) and 4.3 (introduction to IVR system).[17] We concentrate on the first call because menu steps experienced by all users within treatment and within control are identical.[18,19]

Figure 3 displays the probability of treatment and control users not completing a menu step in their first call, conditional on having completed the previous step, across all menu steps in each of the four experiments. In each plot, the menu steps affected by the system modification under study are shaded in gray.

The displayed patterns of conditional attrition across menu steps provide insight into how the tested modifications affected information access. First, the registration experiments show that registration not only generates substantial attrition while users navigate the registration menus, but also changes the composition of users who reach subsequent stages of the system. In the full-registration experiment (1.1), users who survive registration exhibit lower conditional attrition than users assigned to no registration at each subsequent menu. A similar pattern emerges in the short-registration experiment (4.1): although conditional attrition remains higher at the first menu following registration, it is lower from the main menu onward. These patterns are consistent with selection induced by registration. Users with a greater propensity to disengage are disproportionately filtered out during registration. Removing registration therefore allows marginal users who would otherwise have exited the system to progress further toward content.

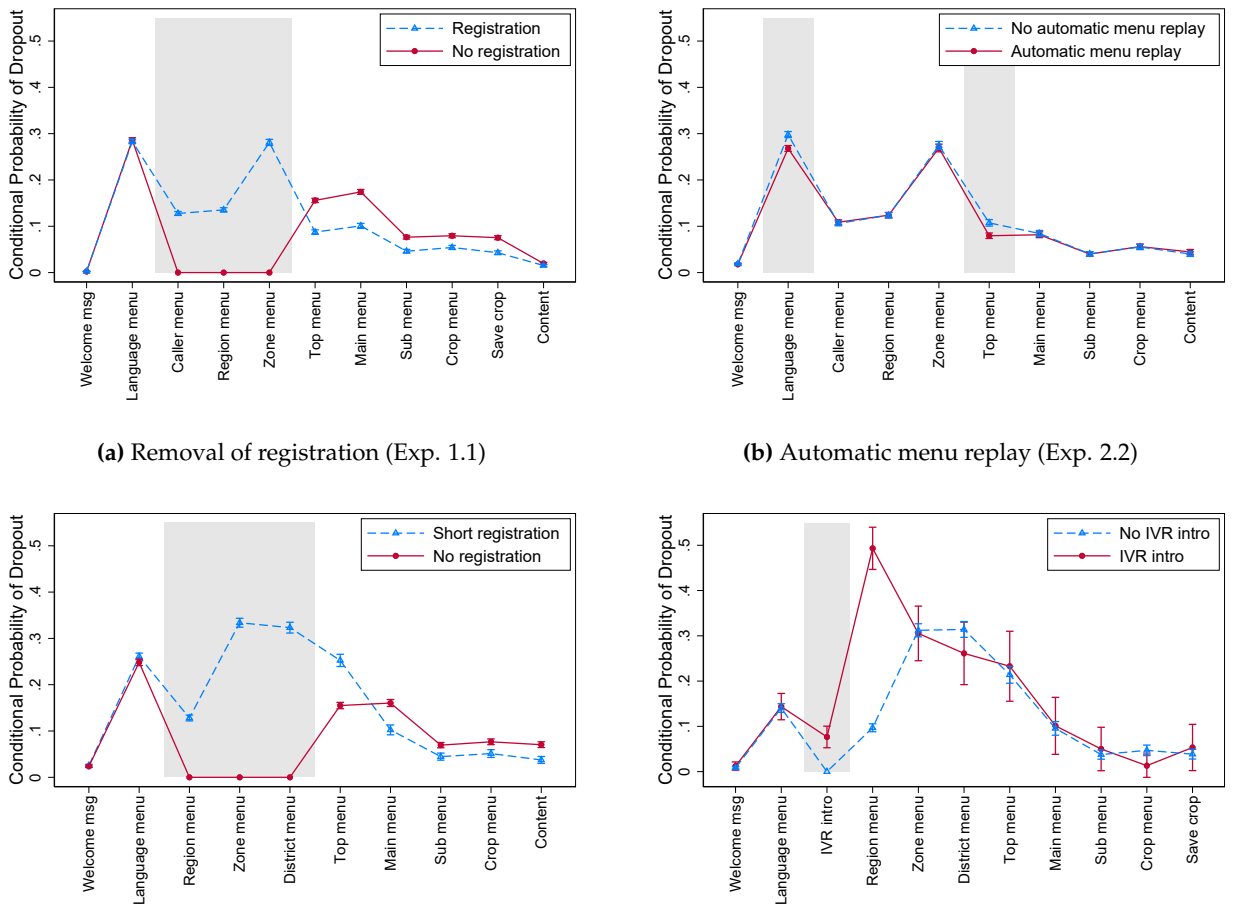
Second, the effects of automatic menu replay are concentrated at the menu steps directly affected by the modification. Automatic replay reduces conditional attrition at these steps, while attrition at subsequent, unmodified steps closely tracks that in the control group. This result indicates that making individual interactions with the system more forgiving can prevent users from dropping out when they fail to respond within the initially allotted time.

Third, the IVR-introduction and the short-registration experiment affect attrition beyond the step at which they are introduced. In the former experiment, treated users face an additional introductory step and remain more likely to drop out at the following region menu. A similar pattern is visible in the short-registration experiment, where users who complete registration initially exhibit higher conditional attrition at the subsequent top menu before displaying lower attrition thereafter. These patterns are consistent with the additional steps imposing costs that carry forward to subsequent interactions. For example, navigating additional steps before reach-

[17] We do not consider experiment 1.4 (removal of prompt to save crop) because the prompt was the last step before content was played, hence there are no interesting dynamics to study. For system modifications that were tested repeatedly, we focus on the first test.

[18] At subsequent calls, this is not necessarily the case because a user who has to register may have provided no registration information on the first call and would thus be directed through all registration questions on their second call while another user in the same treatment group may have (partially) completed registration on their first call and would thus not face all registration questions on their second call.

[19] In the Appendix, we demonstrate that all documented patterns are robust to including all call during the first day (Figure A.7) or first month (Figure A.8) of a user interacting with the system.



(a) Removal of registration (Exp. 1.1)

(b) Automatic menu replay (Exp. 2.2)

(c) Shortened registration (Exp. 4.1)

(d) Introduction to IVR (Exp. 4.3)

Figure 3: Conditional user attrition during first call

This figure displays the conditional probability of a user not completing a menu step conditional on having completed the previous step across all menu steps, separately for four different experiments across panels (a) through (d). In all cases, the sample is restricted to the first call users ever make to the hotline. Further sample restrictions are introduced to hold the menu steps experienced by users within treatment and within control of each experiment constant: In panel (b), the sample is restricted to callers that are assigned to the control group in Experiment 2.1, i.e., callers that are exposed to the registration process. In panel (c), the sample is restricted to callers that are assigned to the control group in Experiment 4.3 or fall outside the sample for this experiment, i.e., callers that are not exposed to the IVR introduction. In panel (d), the sample is restricted to callers that are assigned to the treatment group in Experiment 4.1, i.e., callers that are exposed to the short registration process. Menu steps are listed along the x-axis. Steps affected by the system modification under study are shaded in gray. Markers indicate point estimates and whiskers indicate 95% confidence intervals. For steps that do not exist in a given treatment or control group, the conditional probability of dropout is zero by construction.

ing the desired content may generate fatigue, impatience, or frustration that increases users' propensity to disengage even after those steps have been completed. While the experiments do not distinguish among these mechanisms, the persistence of elevated attrition beyond the modified step suggests that the costs imposed by additional interactions need not be confined to the interactions themselves.

Finally, patterns across experiments suggest that attrition is not driven solely by the duration of calls. Attrition is relatively limited during passive steps, such as the welcome message and IVR introduction, during which users are only required to listen. It is substantially higher at steps requiring users to actively navigate the system, even when these steps are of similar duration to the welcome message, such as the caller type menu, the top menu and the main menu. Attrition is particularly high at longer and more complex menus, such as the zone and district menus, which contain up to 24 and 26 response options, respectively. Moreover, automatic menu replay increased information delivery without eliminating any menu steps and despite marginally lengthening calls for users requiring a replay. Taken together, these patterns suggest that the benefits of reducing friction cannot be explained solely by shorter exposure to the risk of call disconnection. Rather, the demands placed on users as they navigate the system — including the need to process prompts, make choices, provide information, and respond within a limited period — appear to constitute an important barrier to accessing content.

6 Conclusion

This paper develops and applies a data analytics approach to improving digital public service delivery in the context of Ethiopia's digital agricultural advisory system. The approach combines analysis of system-generated data with qualitative research and iterative experimentation to identify service delivery frictions, investigate their underlying causes, and test potential remedies. Funnel analysis was used to identify points of high user attrition, while a user survey and focus group discussions with farmers and other stakeholders helped diagnose the causes of that attrition. Seven potential system improvements were then developed and evaluated through iterative A/B tests, with modifications refined, retested, adopted, or discarded in response to the evidence.

Three simple system modifications — removing the profile registration process, removing the prompt to save a crop, and automatically replaying menus — significantly increased users' eventual access to agricultural advice. The most effective intervention, removal of the multi-step registration process, increased access to agricultural advice by 18%. Importantly, these improvements came at negligible cost in terms of foregone information: the information no longer collected was either highly inaccurate and therefore of limited value for targeting messages to farmers, or already available elsewhere in the system. The remaining four modifications did not increase information delivery.

A common feature of the three successful modifications was that they reduced friction along the user's path to content. Two eliminated requests for information from users, while the third

reduced the consequences of failing to respond to a menu prompt in time. This pattern highlights the potential costs of seemingly minor hassles in digital public services, particularly when users must navigate multiple steps before reaching the service they seek. At the same time, the failure of four plausible modifications to increase information delivery underscores the importance of testing potential improvements rather than assuming that changes intended to improve the user experience will do so. Qualitative research and analysis of system data can generate hypotheses about how services might be improved, but experimentation provides a mechanism for determining which changes actually improve service delivery.

The experience also illustrates the potential for analytical evidence to feed back into government service design. All three modifications shown to improve information delivery were eventually adopted system-wide, generating persistent improvements in system performance. Adoption was not always immediate and, in some cases, followed multiple rounds of testing and consideration of alternative designs. Nevertheless, the government's eventual response suggests that at least some service delivery frictions persist not because policymakers have a strategic interest in maintaining them, but because their consequences and potential remedies are not sufficiently visible. Systematic feedback loops that identify frictions, diagnose their causes, and experimentally evaluate solutions can therefore provide a mechanism for continuous improvement in public service delivery.

Such feedback loops are commonplace in the private technology sector, where firms routinely analyze user behavior and experiment with product modifications, and face strong incentives to respond when frictions cause users to disengage. Governments generally face weaker incentives to retain users and may therefore invest less in systematically identifying and removing such frictions. Even in the setting studied here, the government responded to the evidence generated by the analytical process but did not itself routinely conduct funnel analyses, focus groups, and A/B tests to identify and remove service delivery frictions. This may reflect limited political returns to improving relatively mundane aspects of implementation, constraints on the human capital and analytical capacity required to conduct such work, or both. Data-driven feedback loops cannot replicate the full set of incentives created by market competition, but they may provide one mechanism for bringing continuous diagnosis and improvement into the design of public services.

These constraints also raise a question about who should invest in developing and applying such approaches. To the extent that lessons about how to identify and remove service delivery frictions are transferable across programs and countries, the knowledge generated by this work extends beyond the government or service in which it is produced. This creates a potential rationale for external support for experimentation and analytical capacity, particularly when findings are documented and disseminated so that they can inform service design elsewhere. As governments increasingly deliver services through digital systems, the resulting behavioral data and capacity for rapid experimentation create new opportunities to build such feedback loops into public service delivery. Realizing that potential, however, requires not only digitizing services but

also developing the institutional capacity and incentives to continuously learn from how citizens actually use them.

References

Abate, G. T., Bernard, T., de Brauw, A. & Minot, N. (2018), 'The impact of the use of new technologies on farmers wheat yield in Ethiopia: evidence from a randomized control trial', *Agricultural Economics* **49**(4), 409–421.

Acemoglu, D. & Robinson, J. A. (2012), *Why Nations Fail: The Origins of Power, Prosperity and Poverty*, 1st edn, Crown, New York.

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

Angrist, N., Bergman, P. & Matsheng, M. (2022), 'Experimental evidence on learning using low-tech when school is out', *Nature Human Behaviour* **6**(7), 941–950.
URL: *https://doi.org/10.1038/s41562-022-01381-z*

Angrist, N., Cullen, C. & Magat, J. (2025), Cheaper (and more effective) by the dozen: Evidence from 12 randomized a/b tests optimizing tutoring for scale, Ssrn working paper, Social Science Research Network (SSRN). Available at SSRN: https://ssrn.com/abstract=5755162.
URL: *https://ssrn.com/abstract=5755162*

Athey, S. (2025), 'Presidential address: The economist as designer in the innovation process for socially impactful digital products', *American Economic Review* **115**(4), 1059–1099. Presidential Address.

Athey, S., Cole, S., Nath, S. & Zhu, S. J. (2026), Targeting, personalization, and engagement in an agricultural advisory service, Working Paper 34951, National Bureau of Economic Research. NBER Working Paper No. 34951.
URL: *https://www.nber.org/papers/w34951*

Bandiera, O., Prat, A. & Valletti, T. (2009), 'Active and Passive Waste in Government Spending: Evidence from a Policy Experiment', *American Economic Review* **99**(4), 1278–1308.

Benzarti, Y. (2020), 'How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs', *American Economic Journal: Economic Policy* **12**(4), 38–57.

Bergman, P., Lasky-Fink, J. & Rogers, T. (2020), 'Simplification and defaults affect adoption and impact of technology, but decision makers do not realize it', *Organizational Behavior and Human Decision Processes* **158**, 66–79.

Bettinger, E. P., Long, B. T., Oreopoulos, P. & Sanbonmatsu, L. (2012), 'The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment*', *The Quarterly Journal of Economics* **127**(3), 1205–1242.

Bhargava, S. & Manoli, D. (2015), 'Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment', *American Economic Review* **105**(11), 3489–3529.

Callen, M., Gulzar, S., Hasanain, A., Khan, M. Y. & Rezaee, A. (2020), 'Data and policy decisions: Experimental evidence from pakistan', *Journal of Development Economics* **146**, 102523.

Capital Ethiopia (2026), 'Farming as a business, not survival'.
URL: *https://capitalethiopia.com/2026/02/01/farming-as-a-business-not-survival/*

Casaburi, L., Kremer, M., Mullainathan, S. & Ramrattan, R. (2019), Harnessing ICT to increase agricultural production: Evidence from Kenya. Working Paper.
URL: *https://www.econ.uzh.ch/dam/jcr:873845ce-de4d-4366-ba9a-d60accda577d/SMS*$_{p}$*aper*$_{w}$*ith*$_{t}$*ables*$_{2}$0190923$_{m}$*erged*

Cole, S. A. & Fernando, A. N. (2021), ''mobile'izing agricultural advice: Technology adoption, diffusion and sustainability', *The Economic Journal* **131**(633), 192–219.
URL: *https://academic.oup.com/ej/article-abstract/131/633/192/5867759*

DellaVigna, S., Kim, W. & Linos, E. (2022), Bottlenecks for Evidence Adoption, Working Paper 30144, National Bureau of Economic Research.

Dumais, S., Jeffries, R., Russell, D. M., Tang, D. & Teevan, J. (2014), Understanding User Behavior Through Log Data and Analysis, *in* J. S. Olson & W. A. Kellogg, eds, 'Ways of Knowing in HCI', Springer, New York, NY, pp. 349–372.

Ellis, S. & Brown, M. (2017), *Hacking Growth: How Today's Fastest-Growing Companies Drive Breakout Success*, Random House Audio Assets.

Fabregas, R., Kremer, M., Lowes, M., On, R. & Zane, G. (2025), 'Digital information provision and behavior change: Lessons from six experiments in East Africa', *American Economic Journal: Applied Economics* **17**(1), 527–566.
URL: *https://www.aeaweb.org/articles?id=10.1257/app.20220072*

Fabregas, R., Kremer, M. & Schilbach, F. (2019), 'Realizing the potential of digital development: The case of agricultural advice', *Science* **366**(6471), eaay3038.

Fafchamps, M. & Minten, B. (2012), 'Impact of SMS-based agricultural information on Indian farmers', *The World Bank Economic Review* **26**(3), 383–414.
URL: *http://www.jstor.org/stable/41679567*

Figueiredo Walter, T., Gray-Lobe, G. & Kabay, S. (2023), Can brick phones bridge the digital learning divide? Evidence from SMS-based math practice. EdWorkingPaper No. 23-791. Annenberg Institute for School Reform at Brown University.
URL: *https://files.eric.ed.gov/fulltext/ED672337.pdf*

Finan, F., Olken, B. A. & Pande, R. (2015), The Personnel Economics of the State, Working Paper 21825, National Bureau of Economic Research.

Ganimian, A. J., Muralidharan, K. & Walters, C. R. (2024), 'Augmenting state capacity for child development: Experimental evidence from India', *Journal of Political Economy* **132**(5), 1565–1602.
URL: *https://www.journals.uchicago.edu/doi/full/10.1086/728109*

Gelb, A., Mittal, N. & Mukherjee, A. (2019), Towards Real-Time Governance: Using Digital Feedback to Improve Service, Voice, and Accountability, Working Paper, Center for Global Development.

Hjort, J., Moreira, D., Rao, G. & Santini, J. F. (2021), 'How research affects policy: Experimental evidence from 2,150 brazilian municipalities', *American Economic Review* **111**(5), 144280.

Kasy, M. & Sautmann, A. (2021), 'Adaptive treatment assignment in experiments for policy choice', *Econometrica* **89**(1), 113–132.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA17527*

Kohavi, R., Longbotham, R., Sommerfield, D. & Henne, R. M. (2009), 'Controlled experiments on the web: survey and practical guide', *Data Mining and Knowledge Discovery* **18**(1), 140–181.

Kohavi, R., Tang, D. & Xu, Y. (2020), *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*, Cambridge University Press.

Koning, R., Hasan, S. & Chatterji, A. (2022), 'Experimentation and start-up performance: Evidence from a/b testing', *Management Science* **68**(9), 6434–6453.
URL: *https://doi.org/10.1287/mnsc.2021.4209*

Moynihan, D. P. & Herd, P. (2019), *Administrative Burden: Policymaking by Other Means*, Russell Sage Foundation, New York.

Muralidharan, K., Romero, M. & Wthrich, K. (2019), Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments, Working Paper 26562, National Bureau of Economic Research.

Muralidharan, K. & Singh, A. (2020), Improving Public Sector Management at Scale? Experimental Evidence on School Governance India, Working Paper 28129, National Bureau of Economic Research.

Pahlka, J. (2023), *Recoding America: Why Government Is Failing in the Digital Age and How We Can Do Better*, Metropolitan Books, Henry Holt and Company, New York, NY.

Pritchett, L., Samji, S. & Hammer, J. S. (2013), It's All About MeE: Using Structured Experiential Learning ('e') to Crawl the Design Space, Working Paper 322, Center for Global Development.

Sunstein, C. R. (2022), *Sludge: What Stops Us from Getting Things Done and What to Do about It*, MIT Press, Cambridge, MA.

Tang, D., Agarwal, A., O'Brien, D. & Meyer, M. (2010), Overlapping Experiment Infrastructure: More, Better, Faster Experimentation, *in* 'Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', KDD '10, Association for Computing Machinery, New York, NY, USA, pp. 17–26.

Van Campenhout, B., Spielman, D. J. & Lecoutere, E. (2021), 'Information and communication technologies to provide agricultural advice to smallholder farmers: Experimental evidence from Uganda', *American Journal of Agricultural Economics* **103**(1), 317–337.
URL: *https://onlinelibrary.wiley.com/doi/full/10.1002/ajae.12089*

Viviano, D. & Rudder, J. (2026), Policy design in experiments with unknown interference. Working Paper.
URL: *https://arxiv.org/pdf/2011.08174*

A Appendix

A.1 Examples of content messages

A.1.1 Example 1

> "Currently, two sources of nutrients are recommended: Urea (46% N) and NPS Ammonium Sulfo Phosphate (19% N + 38P2O5+ 7S). The recommendations for nutrient levels vary from place to place and additionally depend on the amount of rainfall. In areas with a long history of cultivation and rainfall above 800mm during the crop season, the optimum level of fertilizer application is most likely in the range of 200 to 300 kg/ha of Urea (92-138 kg of N/ha) and 100 kg/ha of NPS. However, if farmers find these levels to be high and difficult to afford, then 150 kg/ha of Urea and 100 kg/ha of NPS can be applied with the understanding that grain yield will be somewhat reduced but still attractive provided these inputs are applied efficiently. In areas for which no specific fertilizer recommendations are available, 100kg/ha of NPS and 100kg/ha of Urea can be applied."

A.1.2 Example 2

> "Use of appropriate seed rate helps to reduce the seed wastage, provides good crop growth and increases yield. Seeds must be sowed evenly. Choose a variety with an above 90% growth rate and plant 80-100kg/ha. Place seeds evenly in furrows 30cm apart. To make furrows that are 30cm apart use a plow with wooden blades that are 30cm apart. This allows for easy hand weeding, urea side application, and crop management. If using a tractor to plant seeds, do so in furrows 20cm apart and 3-4cm deep."

A.1.3 Example 3

> "Choose a wheat variety that has high yield and resistance to disease, especially rust. Also use varieties that are ideal for your areas agro-ecology. For highlands and areas with sufficient rainfall use varieties like Denda, Alidoro, Digaloo and Hidasse. For medium altitude areas use varieties like Kekebo, King bird, Ogolcho and Shorim. You can also use other similar varieties as recommended by your Development Agent."

A.2 Experimental integrity

At the beginning of each incoming call, the system automatically determines the eligibility of users to participate in the different ongoing experiments based on the definitions of the experimental samples. Then eligible users are randomly assigned to treatment and control in the corresponding

experiment(s). Randomization is independent across experiments (even if experiments are being carried out simultaneously).

To provide evidence of experimental integrity, we test whether the completion of the step in the IVR menu that precedes the step affected by the treatment is uncorrelated with treatment. Table A.3 presents these results for all experiments. In no case, can we reject balance. Two cautionary remarks are necessary, however. First, we cannot provide evidence of balance for two of the ten experiments using this approach because in these cases, the previous step in the IVR menu determines the eligibility for the experiment. In both cases, the previous step is language selection and users have to select Amharic to be eligible. In other words, the previous step is completed by everyone in the experimental sample by design. However, we present an alternative test for these two experiments in Table A.4 where we test whether the time of day at the first call of treatment and control users differs systematically. Reassuringly, the difference is very small and statistically insignificant. Second, for experiments where the treatment includes a modification of the language menu the preceding step in the IVR menu is listening to the welcome message. This step is completed by the vast majority of users (99%), thus leaving little variation to test for balance.

A.3 IVR introduction script

This is an English translation of the script of the introduction to the IVR:

- Welcome! Since this is your first time, I will now explain to you how to use this system to find the information you want.
- Calling the system is free.
- This is an IVR system, where you press numbers to navigate. If you already know how to use such systems, or this one, and dont want to listen to the training, press 9 at any time to skip this training and start using the system now.
- Recorded messages are replayed by a machine. You are not talking to an actual person. The machine cannot hear you and answer to your voice. You can only talk back to the system by pressing the buttons.
- The end goal is to listen to recorded messages with agricultural information. The system has information about different crops and different practices.
- Call 8028. Put the phone on speaker, so you can press the buttons while listening to the messages.
- The system will ask you a question and read several options for you, and you will choose the option you want and press the button with the number of that option.

- When the voice reads these options to you, listen and remember the one that is most relevant. When you hear the option you want, press the button.
- Put the phone on speaker, so you can press the buttons while listening to the messages. For example: You will be asked to choose the crop that you want to listen a content about: For food barley, press 01. For malt barley, press 02. For maize, press 03. For sorghum, press 04. Etc. You press 0 and 3 if you want to listen about Maize.
- There might be multiple relevant options, select one. If you want to listen to another one later, you can call again.
- If the call failed, be patient and try a few times.
- If you are unable to use the system, maybe a child can help. Or a teacher. Or the DA. Thank you.

A.4 Appendix Tables

Table A.1: Research Design

Round	Time period	No.	Experiment	Treatment	Sample	Pr[Treat]	# Arms
1	29 Nov 2017 - 31 Dec 2017	1.1	Registration-1	Remove registration	New users	1/2	1
		1.2	Menu Replay-1	Add menu replay after 10 seconds of inactivity	New users	1/2	1
		1.3	Menu Pauses-1	Add pauses between language options	New users	1/2	1
		1.4	Crop Prompt-1	Remove prompt to save crop	All users	1/2	1
2	28 Feb 2018 - 31 Mar 2018	2.1	Registration-1	Remove registration	New users	1/2	1
		2.2	Menu Replay-2	Add menu replay after 4 seconds of inactivity	New users	1/2	1
3	1 May 2018 - 19 Jun 2018	3.1	Registration-2	Voluntary registration	5% random sample of new Amharic-speaking users	2/3	2
4	1 Jul 2018 - 24 Jul 2018	4.1	Registration-3	Shortened registration	New users	1/2	1
		4.2	Menu Replay-2	Add menu replay after 4 seconds of inactivity	New users	1/2	1
		4.3	Introduction-1	Add IVR introduction	New Amharic-speaking users	1/10	2

Between November 2017 and July 2018, ten experiments were conducted over four rounds. This table provides an overview of these experiments, describing when they were carried out (Time period) as well as what number (No.) and title (Experiment) are used to refer to them in the remainder of the paper. The table also specifies treatment in each experiment (Treatment), the experimental sample (Sample), the probability of treatment conditional on being part of the experimental sample (Pr[Treat]) and the number of treatment arms (# Arms).

Table A.2: Summary statistics

	Sample (1A) Exp. 1.1, 1.2, 1.3		Sample (1B) Exp. 1.4		Sample (2) Exp. 2.1, 2.2		Sample (3) Exp. 3.1		Sample (4A) Exp. 4.1, 4.2		Sample (4B) Exp. 4.3	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Ever accessed any content	0.638	0.481	0.637	0.481	0.604	0.489	0.826	0.379	0.573	0.495	0.696	0.46
Content items accessed	5.204	10.823	5.199	10.834	4.724	9.838	6.145	9.766	4.202	9.426	4.927	10.27
Unique content items accessed	2.951	4.974	2.948	4.972	2.845	4.977	3.943	5.469	2.721	5.014	3.238	5.335
Ever selected a language	0.887	0.316	0.887	0.316	0.855	0.352	1	0	0.854	0.354	1	0
Ever selected a crop	NA	NA	0.646	0.478	NA	NA	NA	NA	NA	NA	NA	NA
System ever terminated call at lang menu	0.043	0.204	NA	NA	0.057	0.232	NA	NA	0.044	0.204	NA	NA
System ever terminated call at top menu	0.103	0.303	NA	NA	0.107	0.309	NA	NA	0.105	0.307	NA	NA
Any call ended after crop selection	NA	NA	0.048	0.213	NA	NA	NA	NA	NA	NA	NA	NA
Ever completed registration	NA	NA	NA	NA	NA	NA	0.053	0.224	NA	NA	NA	NA
Treated in experiment 1.1	0.5	0.5	0.492	0.5	0	0	0	0	0	0	0	0
Treated in experiment 1.2	0.499	0.5	0.492	0.5	0	0	0	0	0	0	0	0
Treated in experiment 1.3	0.499	0.5	0.491	0.5	0	0	0	0	0	0	0	0
Treated in experiment 1.4	0.498	0.5	0.498	0.5	0	0	0	0	0	0	0	0
Treated in experiment 2.1	0	0	0	0	0.499	0.5	0	0	0	0	0	0
Treated in experiment 2.2	0	0	0	0	0.5	0.5	0	0	0	0	0	0
Treated in experiment 3.1	0	0	0	0	0	0	0.658	0.475	0	0	0	0
Treated in experiment 4.1	0	0	0	0	0	0	0	0	0.502	0.5	0.502	0.5
Treated in experiment 4.2	0	0	0	0	0	0	0	0	0.491	0.5	0.502	0.5
Treated in experiment 4.3	0	0	0	0	0	0	0	0	0.039	0.195	0.099	0.299
N	67316		68377		77628		1072		30517		11370	

This table provides summary statistics for the seven samples of callers used to estimate the treatment effects in all ten experiments. The variables are defined as follows. *Ever accessed any content* is an indicator variable that takes value one if the caller accessed at least one content message in the first two months following their initial call, and zero otherwise. *Content items accessed* indicates the number of content items accessed by the caller in the first two months following their initial call. If a caller accesses the same content item multiple times, they are counted repeatedly. *Unique content items accessed* indicates the number of unique content items accessed by the caller in the first two months following their initial call. *Ever selected a language* is an indicator variable that takes value one if the caller ever selected a language in the first two months following their initial call, and zero otherwise. *Ever selected a crop* is an indicator variable that takes value one if the caller ever selected a crop in the first two months following their initial call, and zero otherwise. *System ever terminated call at lang menu* is an indicator variable that takes value one if any call of the caller in the first two months following their initial call was terminated by the system upon reaching the language menu, and zero otherwise (the system terminates calls automatically if the caller remains inactive). *System ever terminated call at top menu* is an indicator variable that takes value one if any call of the caller in the first two months following their initial call was terminated by the system upon reaching the top menu, and zero otherwise. *Any call ended after crop selection* is an indicator variable that takes value one if any call of the caller in the first two months following their initial call ended (either terminated by the system or hung up by the caller) after selecting a crop but before the content message was played, and zero otherwise. *Ever completed registration* is an indicator variable that takes value one if the caller completed the registration process in the first two months following their initial call, and zero otherwise. *Treated in experiment i* is an indicator variable that take value one if the caller was treated in experiment *i*, and zero otherwise. *N* indicates the number of callers in each sample. Some variables are only used as outcomes for specific treatments in order to better understand the underlying mechanisms. Therefore, they are not available in all samples, as indicated by *NA*.

Table A.3: Balance: Previous step completed

Panel A: Round 1

Experiment	1.1	1.2	1.3	1.4
Sample	(1A)	(1A)	(1A)	(1B)
Previous step	Language selected	Welcome msg	Welcome msg	Crop selected
Treated	0.002	0.000	0.000	-0.002
	(0.002)	(0.000)	(0.000)	(0.004)
R2	0.000	0.000	0.000	0.000
N	67316	67316	67316	68377
Control mean	0.886	0.999	0.999	0.647

Panel B: Round 2

Experiment	2.1	2.2
Sample	(2)	(2)
Previous step	Language selected	Welcome msg
Treated	-0.000	0.000
	(0.003)	(0.001)
R2	0.000	0.000
N	77628	77628
Control mean	0.855	0.991

Panel C: Round 3

Experiment	3.1
Sample	(3)
Previous step	Not available

Panel D: Round 4

Experiment	4.1	4.2	4.3
Sample	(4A)	(4A)	(4B)
Previous step	Language selected	Welcome msg	Not available
Treated	-0.003	-0.001	
	(0.004)	(0.001)	
R2	0.000	0.000	
N	30517	30517	
Control mean	0.855	0.984	

See Table A.2 notes for variable definitions. Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Balance: Time of day at first call

	Seconds since midnight	
	(1)	(2)
	Exp. 3.1	Exp. 4.3
Treated	-721	-247
	(1070)	(546)
R2	0	0
N	1072	11370
Control mean	48992	48145

Dependent variable is the number of seconds at the start of the first call that have passed since midnight of the same day. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Exp. 3.1 - Treatment Effects by Arm

	Content accessed			Completed ...
	(1)	(2)	(3)	(4)
	Any	# Items	# Unique Items	Registration
Treatment arm 1	-0.038	-0.223	-0.109	-0.015
	(0.028)	(0.729)	(0.408)	(0.017)
Treatment arm 2	-0.001	-0.836	-0.594	0.002
	(0.028)	(0.728)	(0.408)	(0.017)
R2	0.002	0.001	0.002	0.001
N	1072	1072	1072	1072
Control mean	0.839	6.493	4.174	0.057

See Table A.2 notes for variable definitions. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Exp 4.3 - Treatment Effects by Arm

	Content accessed		
	(1)	(2)	(3)
	Any	# Items	# Unique Items
Treatment arm 1	-0.104***	-0.430	-0.280
	(0.0195)	(0.437)	(0.227)
Treatment arm 2	-0.125***	-0.492	-0.369
	(0.0202)	(0.451)	(0.235)
R2	0.00559	0.000180	0.000334
N	11370	11370	11370
Control mean	0.707	4.973	3.270

See Table A.2 notes for variable definitions. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.5 Appendix Figures

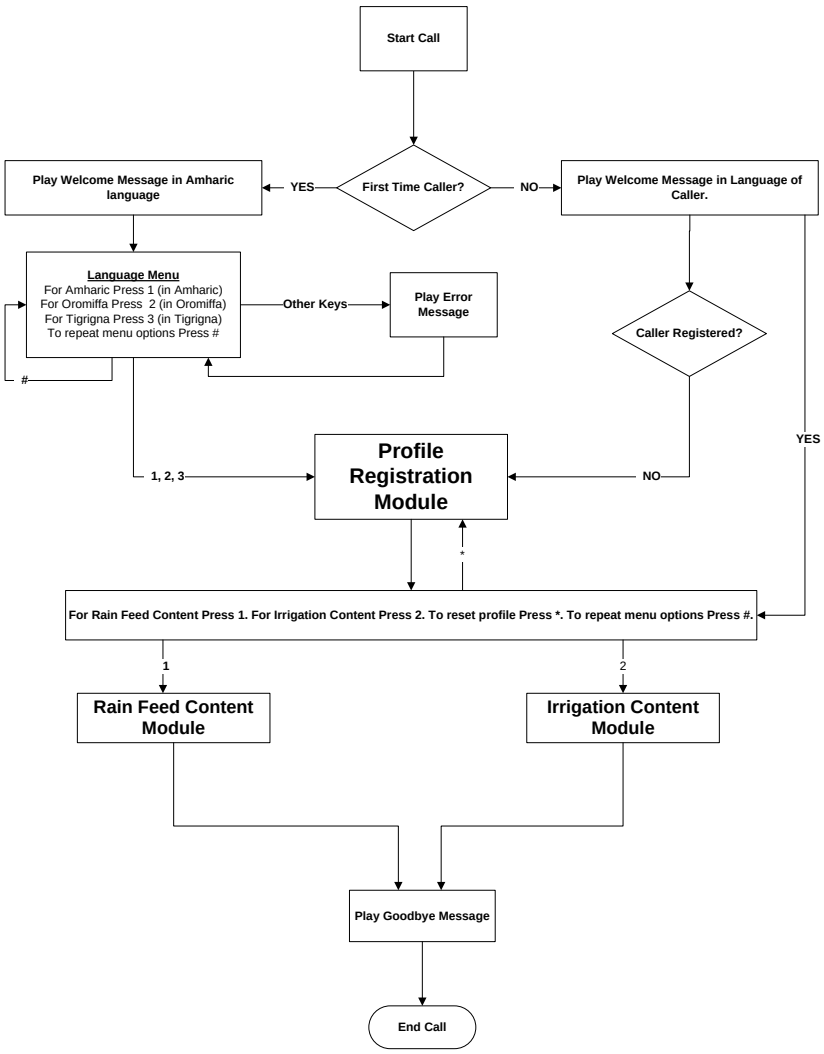


Figure A.1: Flow chart of call to IVR System

19138234	17366	2018-03-26 09:12:16	Incoming call started	162	fa9e5036-ed3c-431
19138247	17366	2018-03-26 09:12:24	Welcome message played	1	fa9e5036-ed3c-431
19138401	17366	2018-03-26 09:13:24	The application hung up the phone	33	fa9e5036-ed3c-431
19138412	17366	2018-03-26 09:13:29	Incoming call started	162	b9db961d-74e6-4f2
19138430	17366	2018-03-26 09:13:37	Welcome message played	1	b9db961d-74e6-4f2
19138463	17366	2018-03-26 09:13:52	Woreda Menu - Enemor Ener	-1	b9db961d-74e6-4f2
19138486	17366	2018-03-26 09:14:01	Gender Menu - Female option selected	128	b9db961d-74e6-4f2
19138487	17366	2018-03-26 09:14:01	Profile registration complete	137	b9db961d-74e6-4f2
19138525	17366	2018-03-26 09:14:17	Top Menu - Rain option selected	301	b9db961d-74e6-4f2
19138534	17366	2018-03-26 09:14:22	Main Menu - Pre Planting option selected	15	b9db961d-74e6-4f2
19138550	17366	2018-03-26 09:14:28	Menu 1 - Land Preparation option selected	280	b9db961d-74e6-4f2
19138566	17366	2018-03-26 09:14:38	Crop Menu - Maize option selected	35	b9db961d-74e6-4f2
19138634	17366	2018-03-26 09:15:11	Content Played - Land Preparation - Maize - Am…	1000506	b9db961d-74e6-4f2
19138654	17366	2018-03-26 09:15:18	Content Menu - Content information for a diffe…	50	b9db961d-74e6-4f2
19138730	17366	2018-03-26 09:16:04	Crop Menu - Go to Main Menu option selected	42	b9db961d-74e6-4f2
19138753	17366	2018-03-26 09:16:17	Top Menu - Reset Profile option selected	303	b9db961d-74e6-4f2
19138755	17366	2018-03-26 09:16:18	Reset Menu - Reset Menu replayed	352	b9db961d-74e6-4f2
19138766	17366	2018-03-26 09:16:24	Top Menu - HHI option selected	302	b9db961d-74e6-4f2
19138778	17366	2018-03-26 09:16:29	HHI Main Menu - Pre Planting option selected	305	b9db961d-74e6-4f2
19138784	17366	2018-03-26 09:16:32	HHIMenu 1 - HHI Pre-irrigation option selected	325	b9db961d-74e6-4f2

Figure A.2: IVR System Call Log

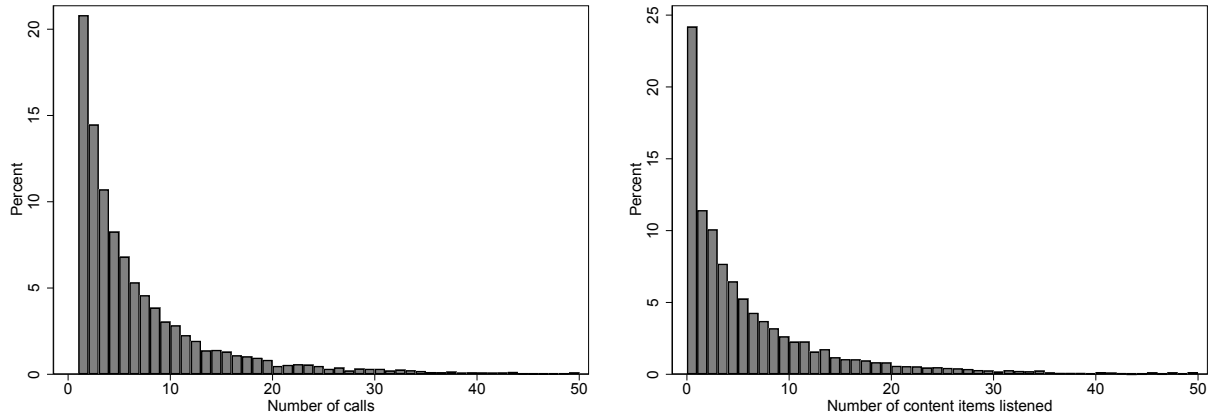


Figure A.3: Distribution of system interactions post registration (two months) across users

Notes: The left panel excludes 78 users that make more than 50 calls over the two months following the completion of the first part of registration and the the right panel excludes 118 users that listen to more than 50 content items over this time period.

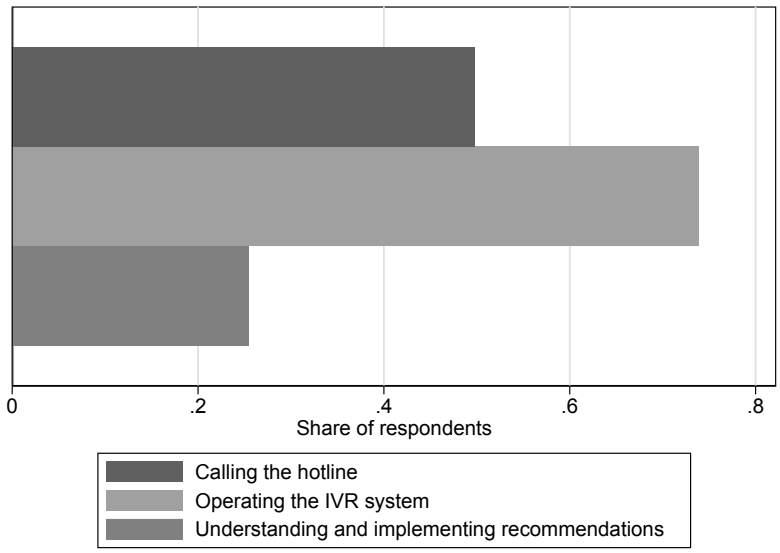


Figure A.4: Barriers to IVR system use as reported by respondents of telephone survey

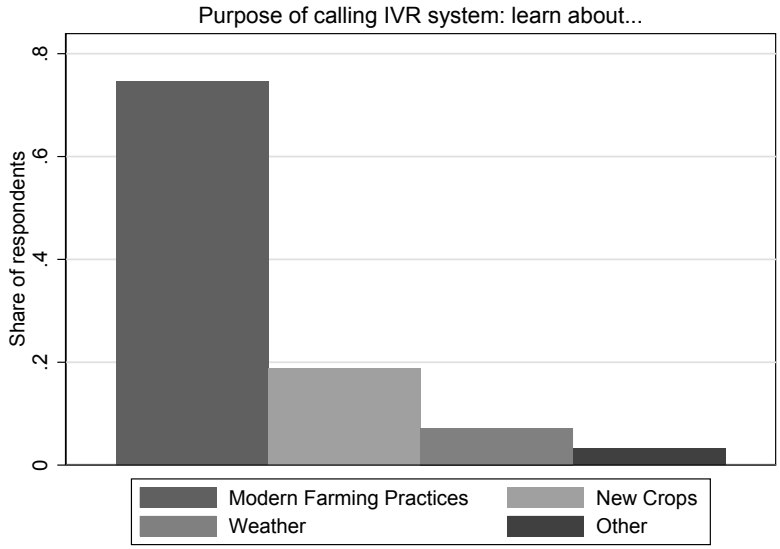


Figure A.5: Reasons for calling the IVR system as reported by respondents of telephone survey

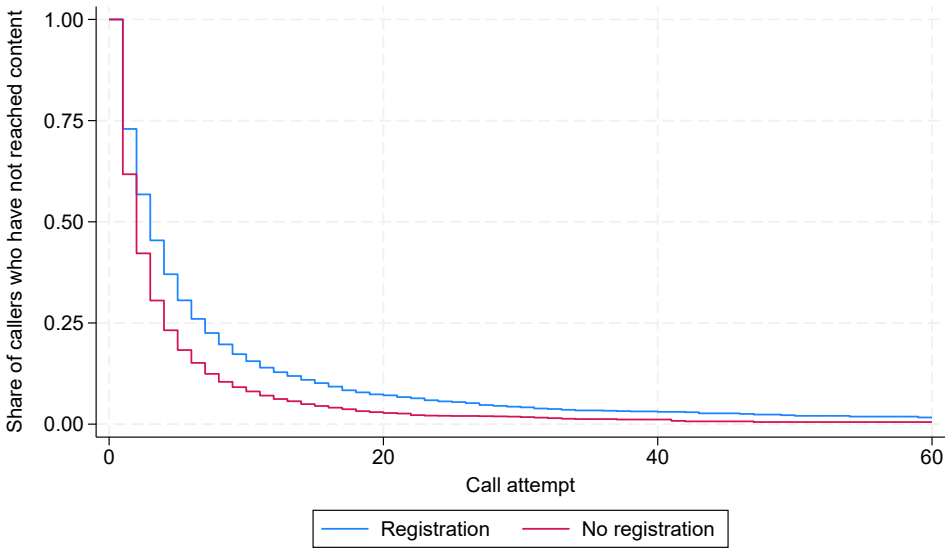
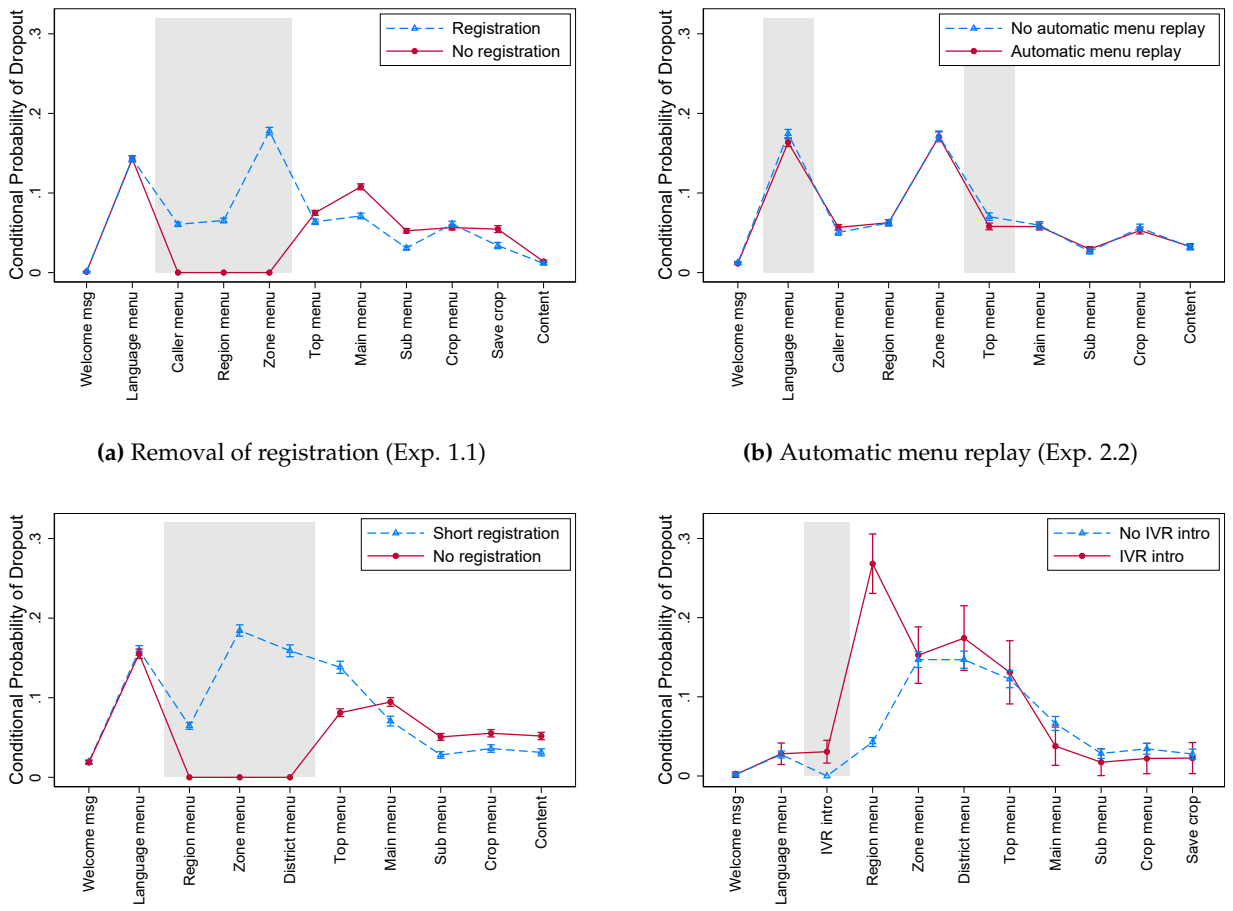


Figure A.6: Content access across calls in Experiment 1.1



(a) Removal of registration (Exp. 1.1)

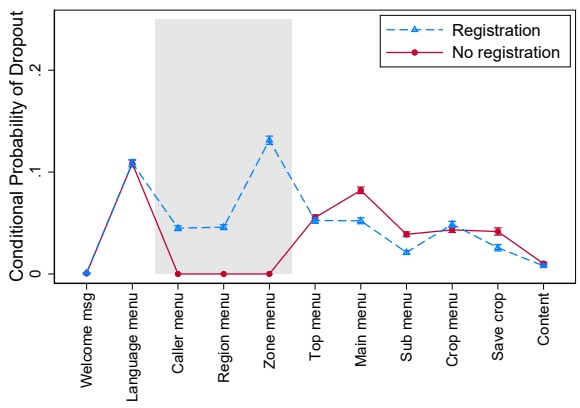
(b) Automatic menu replay (Exp. 2.2)

(c) Shortened registration (Exp. 4.1)

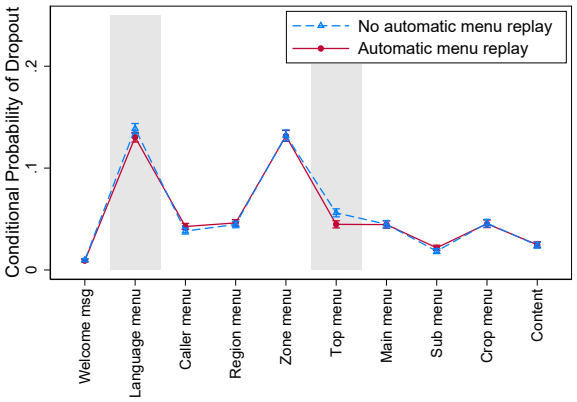
(d) Introduction to IVR (Exp. 4.3)

Figure A.7: Conditional user attrition during first day

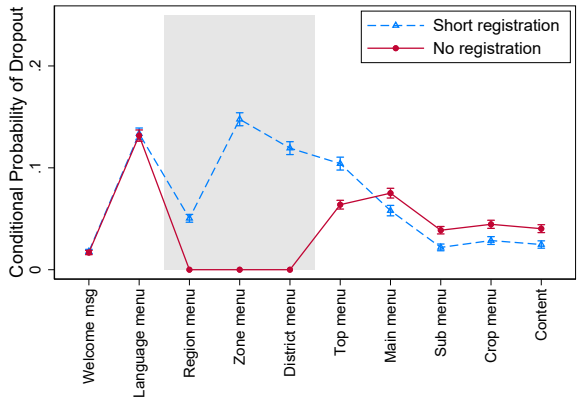
This figure displays the conditional probability of a user not completing a menu step conditional on having completed the previous step across all menu steps, separately for four different experiments across panels (a) through (d). In all cases, the sample is restricted to the day on which users make their first to the hotline. Further sample restrictions are introduced to hold the menu steps experienced by users within treatment and within control of each experiment constant: In panel (b), the sample is restricted to callers that are assigned to the control group in Experiment 2.1, i.e., callers that are exposed to the registration process. In panel (c), the sample is restricted to callers that are assigned to the control group in Experiment 4.3 or fall outside the sample for this experiment, i.e., callers that are not exposed to the IVR introduction. In panel (d), the sample is restricted to callers that are assigned to the treatment group in Experiment 4.1, i.e., callers that are exposed to the short registration process. Menu steps are listed along the x-axis. Steps affected by the system modification under study are shaded in gray. Markers indicate point estimates and whiskers indicate 95% confidence intervals. For steps that do not exist in a given treatment or control group, the conditional probability of dropout is zero by construction.



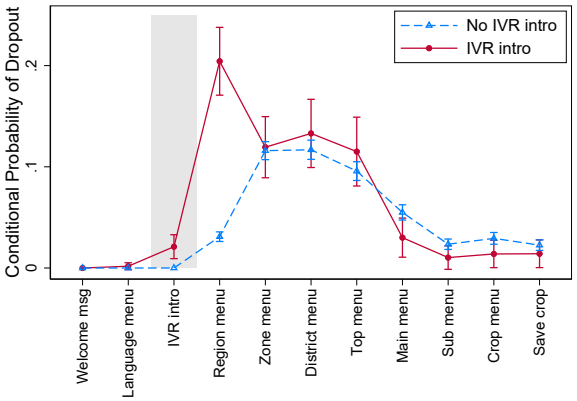
(a) Removal of registration (Exp. 1.1)



(b) Automatic menu replay (Exp. 2.2)



(c) Shortened registration (Exp. 4.1)



(d) Introduction to IVR (Exp. 4.3)

Figure A.8: Conditional user attrition during first month

This figure displays the conditional probability of a user not completing a menu step conditional on having completed the previous step across all menu steps, separately for four different experiments across panels (a) through (d). In all cases, the sample is restricted to the 30 days following the first call of each user to the system. Further sample restrictions are introduced to hold the menu steps experienced by users within treatment and within control of each experiment constant: In panel (b), the sample is restricted to callers that are assigned to the control group in Experiment 2.1, i.e., callers that are exposed to the registration process. In panel (c), the sample is restricted to callers that are assigned to the control group in Experiment 4.3 or fall outside the sample for this experiment, i.e., callers that are not exposed to the IVR introduction. In panel (d), the sample is restricted to callers that are assigned to the treatment group in Experiment 4.1, i.e., callers that are exposed to the short registration process. Menu steps are listed along the x-axis. Steps affected by the system modification under study are shaded in gray. Markers indicate point estimates and whiskers indicate 95% confidence intervals. For steps that do not exist in a given treatment or control group, the conditional probability of dropout is zero by construction.

B Online Appendix

B.1 Long regressions

Table B.1: Round 1 - Sample (1A) - Long Regression

	Content accessed		
	(1) Any	(2) # Items	(3) # Unique Items
T1=1, T2=0, T3=0, T4=0	0.107***	0.798***	0.458***
	(0.010)	(0.233)	(0.107)
T1=0, T2=1, T3=0, T4=0	0.008	0.096	-0.012
	(0.010)	(0.234)	(0.108)
T1=0, T2=0, T3=1, T4=0	-0.001	-0.065	-0.075
	(0.010)	(0.234)	(0.108)
T1=0, T2=0, T3=0, T4=1	0.010	-0.059	0.003
	(0.010)	(0.235)	(0.108)
T1=1, T2=1, T3=0, T4=0	0.113***	0.737***	0.441***
	(0.010)	(0.235)	(0.108)
T1=1, T2=0, T3=1, T4=0	0.107***	0.645***	0.376***
	(0.010)	(0.237)	(0.109)
T1=1, T2=0, T3=0, T4=1	0.114***	1.330***	0.752***
	(0.010)	(0.235)	(0.108)
T1=0, T2=1, T3=0, T4=1	0.031***	0.328	0.185*
	(0.010)	(0.235)	(0.108)
T1=0, T2=1, T3=1, T4=0	0.013	-0.007	-0.021
	(0.010)	(0.235)	(0.108)
T1=0, T2=0, T3=1, T4=1	0.007	0.515**	0.227**
	(0.010)	(0.235)	(0.108)
T1=1, T2=1, T3=1, T4=0	0.107***	0.687***	0.444***
	(0.010)	(0.233)	(0.107)
T1=1, T2=1, T3=0, T4=1	0.114***	1.054***	0.685***
	(0.010)	(0.235)	(0.108)
T1=1, T2=0, T3=1, T4=1	0.117***	0.804***	0.520***
	(0.010)	(0.234)	(0.107)
T1=0, T2=1, T3=1, T4=1	0.010	0.039	0.055
	(0.010)	(0.235)	(0.108)
T1=1, T2=1, T3=1, T4=1	0.119***	1.019***	0.538***
	(0.010)	(0.235)	(0.108)
R2	0.012	0.002	0.003
N	67316	67316	67316
Control mean	0.577	4.710	2.666

Ti is an indicator variable that takes value one if the caller was treated in experiment *i*. See Table A.2 notes for outcome variable definitions. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.2: Round 1 - Sample (1B) - Long Regression

	Content accessed		
	(1)	(2)	(3)
	Any	# Items	# Unique Items
T1=1, T2=0, T3=0, T4=0	0.104***	0.756***	0.448***
	(0.010)	(0.227)	(0.104)
T1=0, T2=1, T3=0, T4=0	0.006	0.062	-0.019
	(0.010)	(0.228)	(0.105)
T1=0, T2=0, T3=1, T4=0	-0.003	-0.102	-0.084
	(0.010)	(0.228)	(0.105)
T1=0, T2=0, T3=0, T4=1	0.009	-0.090	-0.002
	(0.010)	(0.222)	(0.102)
T1=1, T2=1, T3=0, T4=0	0.111***	0.701***	0.433***
	(0.010)	(0.228)	(0.105)
T1=1, T2=0, T3=1, T4=0	0.104***	0.608***	0.367***
	(0.010)	(0.230)	(0.106)
T1=1, T2=0, T3=0, T4=1	0.111***	1.292***	0.742***
	(0.010)	(0.229)	(0.105)
T1=0, T2=1, T3=0, T4=1	0.029***	0.293	0.177*
	(0.010)	(0.229)	(0.105)
T1=0, T2=1, T3=1, T4=0	0.010	-0.042	-0.029
	(0.010)	(0.229)	(0.105)
T1=0, T2=0, T3=1, T4=1	0.005	0.496**	0.225**
	(0.010)	(0.228)	(0.105)
T1=1, T2=1, T3=1, T4=0	0.104***	0.653***	0.436***
	(0.010)	(0.227)	(0.104)
T1=1, T2=1, T3=0, T4=1	0.113***	1.018***	0.677***
	(0.010)	(0.228)	(0.105)
T1=1, T2=0, T3=1, T4=1	0.115***	0.768***	0.511***
	(0.010)	(0.228)	(0.104)
T1=0, T2=1, T3=1, T4=1	0.008	0.001	0.045
	(0.010)	(0.228)	(0.105)
T1=1, T2=1, T3=1, T4=1	0.117***	0.984***	0.529***
	(0.010)	(0.229)	(0.105)
R2	0.011	0.002	0.003
N	68377	68377	68377
Control mean	0.579	4.745	2.674

Ti is an indicator variable that takes value one if the caller was treated in experiment *i*. See Table A.2 notes for outcome definitions.
Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.3: Round 2 - Sample (2) - Long Regression

	Content accessed		
	(1) Any	(2) # Items	(3) # Unique Items
T1=1, T2=0	0.050***	0.452***	0.294***
	(0.005)	(0.100)	(0.051)
T1=0, T2=1	0.008*	0.192*	0.099**
	(0.005)	(0.100)	(0.050)
T1=1, T2=1	0.062***	0.493***	0.339***
	(0.005)	(0.100)	(0.051)
R2	0.003	0.000	0.001
N	77628	77628	77628
Control mean	0.574	4.440	2.662

Ti is an indicator variable that takes value one if the caller was treated in experiment *i*. See Table A.2 notes for outcome definitions. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.4: Round 4 - Sample (4A) - Long Regression

	Content accessed		
	(1) Any	(2) # Items	(3) # Unique Items
T1=1, T2=0, T3=0	-0.170***	-1.017***	-0.658***
	(0.008)	(0.154)	(0.082)
T1=0, T2=1, T3=0	0.015*	-0.090	-0.013
	(0.008)	(0.156)	(0.083)
T1=0, T2=0, T3=1	0.031	0.732	0.657**
	(0.029)	(0.561)	(0.298)
T1=1, T2=1, T3=0	-0.161***	-0.966***	-0.562***
	(0.008)	(0.156)	(0.083)
T1=1, T2=0, T3=1	-0.088***	0.047	0.145
	(0.029)	(0.556)	(0.296)
T1=0, T2=1, T3=1	0.038	0.312	0.201
	(0.028)	(0.547)	(0.291)
T1=1, T2=1, T3=1	-0.150***	-0.357	-0.360
	(0.028)	(0.550)	(0.292)
R2	0.031	0.003	0.004
N	30517	30517	30517
Control mean	0.651	4.695	3.012

Ti is an indicator variable that takes value one if the caller was treated in experiment *i*. See Table A.2 notes for outcome definitions. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.5: Round 4 - Sample (4B) - Long Regression

	Content accessed		
	(1)	(2)	(3)
	Any	# Items	# Unique Items
T1=1, T2=0, T3=0	-0.201***	-1.257***	-0.774***
	(0.013)	(0.287)	(0.149)
T1=0, T2=1, T3=0	0.010	-0.122	0.045
	(0.013)	(0.287)	(0.149)
T1=0, T2=0, T3=1	-0.129***	-0.589	-0.132
	(0.028)	(0.651)	(0.338)
T1=1, T2=1, T3=0	-0.188***	-1.103***	-0.577***
	(0.013)	(0.288)	(0.150)
T1=1, T2=0, T3=1	-0.260***	-1.142*	-0.643*
	(0.028)	(0.644)	(0.334)
T1=0, T2=1, T3=1	-0.124***	-0.914	-0.590*
	(0.028)	(0.643)	(0.334)
T1=1, T2=1, T3=1	-0.321***	-1.662***	-1.210***
	(0.028)	(0.637)	(0.331)
R2	0.052	0.003	0.005
N	11370	11370	11370
Control mean	0.802	5.596	3.597

Ti is an indicator variable that takes value one if the caller was treated in experiment *i*. See Table A.2 notes for outcome definitions.
Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.2 IVR menu flowcharts

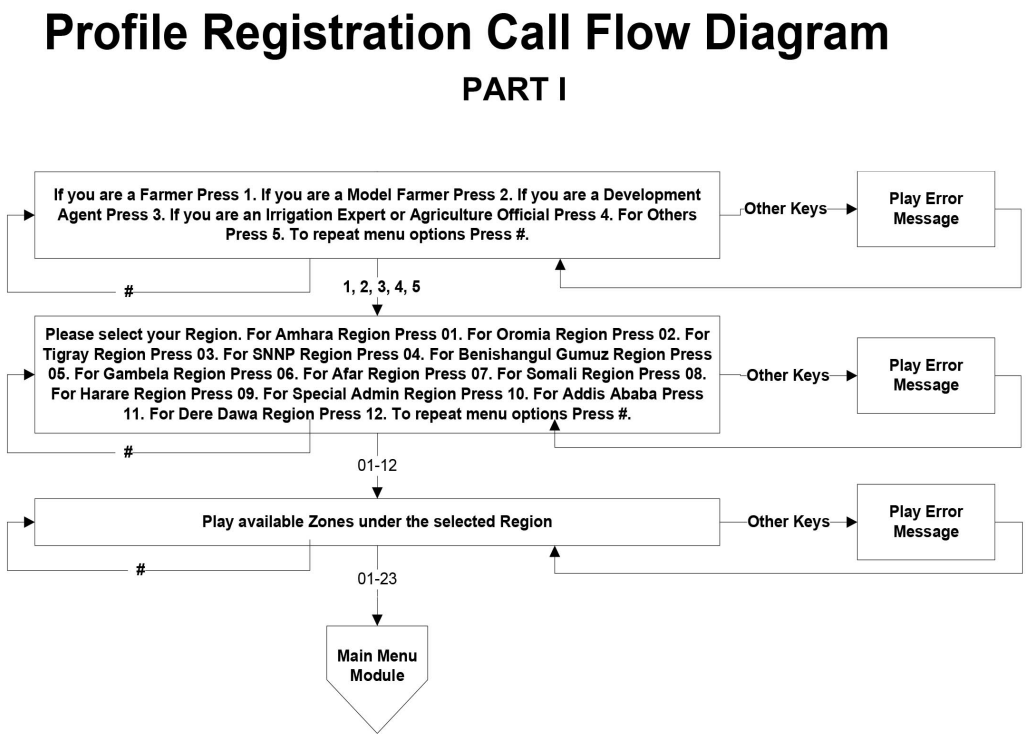


Figure B.1: Flow chart of profile registration module - part 1

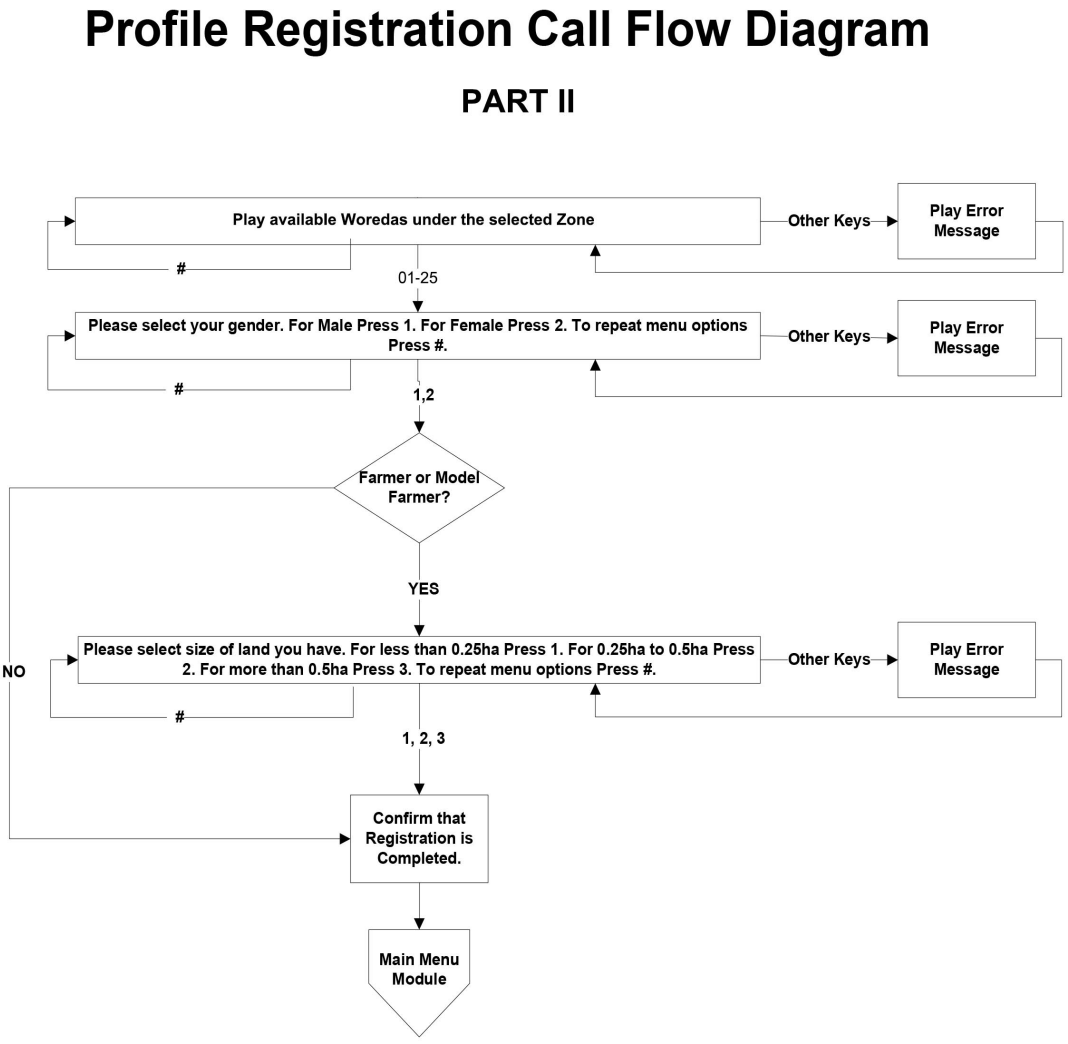


Figure B.2: Flow chart of profile registration module - part 2

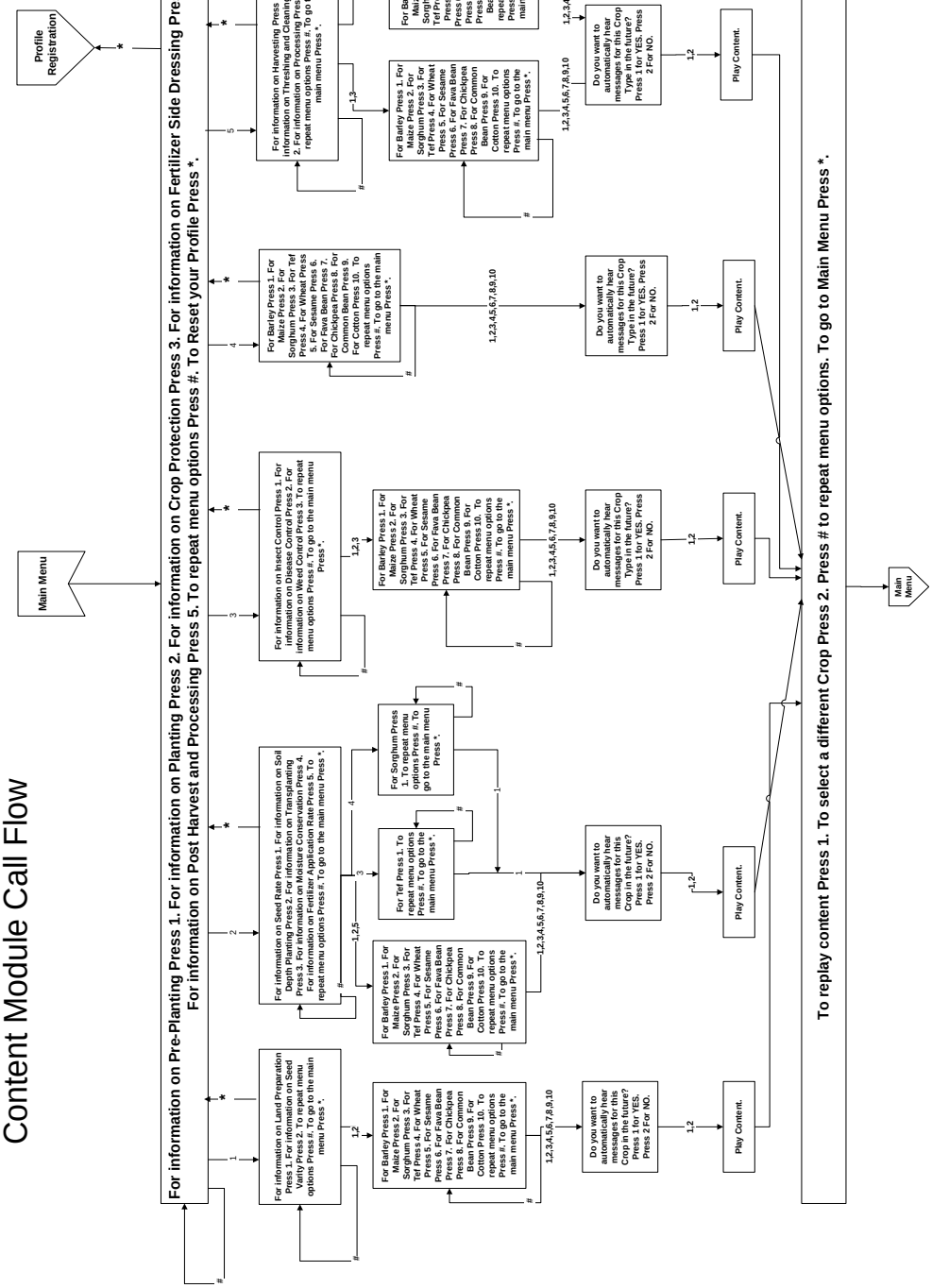


Figure B.3: Flow chart of rainfed content module

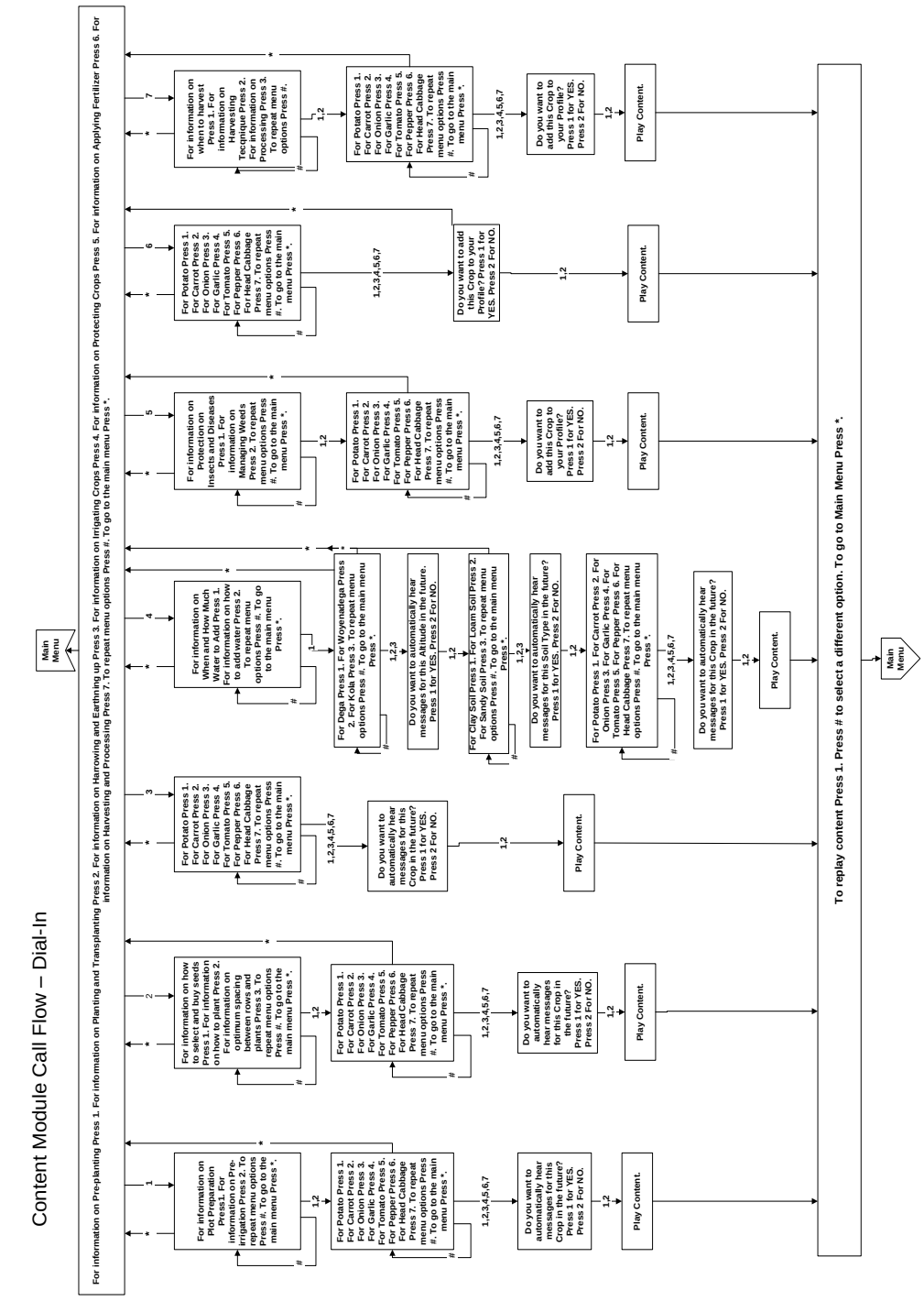


Figure B.4: Flow chart of irrigation content module

B.3 Telephone survey

B.3.1 Sampling

Sampling was stratified by both language and IVR system usage, binned based on the number of unique content items accessed in their most active year, as detailed in Table B.6. Cutoffs between sampling bins were determined based on the distribution of the number of unique contents items accessed in the most active year of each caller.

Table B.6: Telephone survey - Sample stratification by IVR system usage bins

Usage bin	1	2	3	4	5
Number of unique contents	> 20	13-19	7-12	2-6	< 2
Percentile	> 95th	90-95th	75-90th	50-75th	< 50th

B.3.2 Survey instrument

This appendix section displays the complete telephone survey instrument. The survey consisted of six sections:

A Personal information

B Consent

C Demographics and literacy

D Farming practices

E Experience with IVR system

F Experience with alert system

In this paper, we largely focus on the results from section E - Experience with IVR system.

TELEPHONE SURVEY INSTRUMENT

SECTION A: PERSONAL INFORMATION AND CALL STATUS

FO: Unless instructed otherwise, please do not read any of the below questions to the respondent.		
A1.	Date	|___|___| - |___|___| - |___|___|___|___| Day Month Year
A2.	Start Time	|___|___| : |___|___|
A3.	Field Officer (FO) Name	{SELECT FROM LIST}
A4.	Please enter Respondent Phone Number (from checklist)	
A4A.	Please re-enter Respondent's Phone	[error message if different than A4.]
P1.	Respondent's Language	{prefilled}
P2.	Respondent's Region	{prefilled}
P3.	Respondent's Zone	{prefilled}
P4.	Respondent's Woreda	{prefilled}
P5.	Respondent's Gender	{prefilled}
P6.	Type of respondent (Farmers/ Model Farmer / DA / Expert)	{prefilled}
P7	Caller ID	{prefilled}
A5.	Do phone, region, zone, woreda and phone number match checklist?	1 [ ] Yes 0 [ ] No → Please go back and re-enter Farmer's phone number.
A6.	What is the phone status?	1 [ ] Call answered →Continue with the survey Not answered: → **SECTION Z – question Z3 END)** 2 [ ] Phone Off 3 [ ] Not Answering 4 [ ] Busy 5 [ ] Wrong Number 6 [ ] No Longer in Service 7 [ ] Out of Service
A7.	My name is _____ [Check for language understanding and preferences] I am calling on behalf of the Agriculture Transformation Agency (ATA). May I talk to the main user of this phone?	1 [ ] Yes 0 [ ] No -> **A7.A** When could I talk to the main user of this phone? (specify) 1 [ ] Now 2 [ ] Later (specify A7.B) 1 [ ] Date (calendar): 2 [ ] Time: [ ] 1. Between 8am – 11am [ ] 2. Between 11am – 2pm

		[ ] 3. Between 2pm – 5pm 3 [ ] Phone number belongs to a respondent who doesn't speak the language of the enumeration. Specify language and skip to SECTION Z (END) 1. [ ] Amharic 2. [ ] Oromiffa 3. [ ] Tigrigna 4. [ ] Sidamigna 5. [ ] Wolatigna 6. [ ] Agewigna 7. [ ] Kunamigna 8. [ ] Erobigna
A8.	Currently, we are speaking to different users of the free 8028 hotline that offers access to relevant information to improve their farming practices, Are you a Development Agent or Woreda Expert?	1. [ ] Yes **-> FO read: Thank you, at this moment we are contacting just farmers, but in a couple of days one of my colleagues might call you back - Thank you for your time-> END survey section Z** 0. [ ] No
A9.	Our records show that you have been accessing the recommendations of the 8028 hotline on regular and we would like to ask you some questions about your experience with the 8028 Hotline and we could improve it moving forward for other farmers in your woreda. Do you have 20 minutes to answer some questions. [If no, schedule call-back time.]	1 [ ] Now SECTION B (CONSENT) 2 [ ] Later (specify A7.A) 1 [ ] Date (calendar): 2 [ ] Time: 1. [ ] Between 8am – 11am 2. [ ] Between 11am – 2pm 3. [ ] Between 2pm – 5pm SECTION Z – question Z3 (END) 3 [ ] Hang Up → **SECTION Z (END**

SECTION B: CONSENT

FO: Please read.
Precision Agriculture for Development is conducting an evaluation study on the 8028 Program on behalf of Agriculture Transformation Agency (ATA). Busara is contracted by Precision Agriculture for Development to undertake the field data collection. Currently, we are speaking to different farmers to understand the awareness, usage, and benefit of different programs offered by the Agricultural Transformation Agency's (ATA). Particularly, we would like to ask you some questions about the toll free number that farmers can call to access relevant information to improve their farming practices, There are no right or wrong answers to the questions we will be asking you. We strongly encourage you to be open and give us your honest opinions, as well as tell us about your experiences truthfully. We would also like to assure you that what you tell us today will not be shared with others. No experts, or any other who works in the development of the IVR system, will be able to know what any individual says.

The information you give us today will be very helpful 'to improve the quality of the 8028 service being provided to you and millions of other farmers. If you have any questions about the study please contact ATA		
B1.	Do you consent to participate in this survey today?	1. [ ] Yes → **B3** 0. [ ] No
B2.	*[FO: Do not read]* What is the reason why the respondent does not want to participate in the study?	1. [ ] Doesn't have time 2. [ ] Doesn't know/trust ATA 3. [ ] Other:_____
B3.	Did the person hang up after consenting to the survey? *[FO: If the respondent does not have time to complete the phone survey now select "Yes" and be sure to indicate the call-back time in the comments section at the end.]*	1. [ ] Yes → **SECTION Z (END)** 0. [ ] No

SECTION C: DEMOGRAPHICS AND LITERACY

FO read: Now I would like to ask you some questions about your background and the characteristics of your household.		
C1.	(FO: Don't read): What is the respondent's gender?	0. [ ] Male 1. [ ] Female
C2.	Are you a Model Farmer of your village?	1. [ ] Yes 0. [ ] No
C3.	What is your age in years? (FO: If the respondent is unsure, please write down his or her best guess).	|__|__| years *[restrict 18<x<100]*
C4.	In which Region do you live?	{SELECT FROM LIST}
C5.	In which ZONE do you live?	{SELECT FROM LIST}
C6.	In which Woreda do you live?	{SELECT FROM LIST}
C7.	What language do you most frequently use at home? *[FO: Only mark one language]*	1. [ ] Amharic 2. [ ] Oromiffa 3. [ ] Tigrigna 4. [ ] Sidamigna 5. [ ] Wolaytigna 6. [ ]. Agewigna 7. [ ] Kunamigna 8. [ ] Erobigna 99.[ ] Other:_____

C8.	Which other languages do you speak fluently besides the one you just mentioned? *[FO: mark all that apply]*	1. [ ] Amharic 2. [ ] Oromiffa 3. [ ] Tigrigna 4. [ ] Sidamigna 5. [ ] Wolaytigna 6. [ ]. Agewigna 7. [ ] Kunamigna 8. [ ] Erobigna 99 [ ] Other:_____ -97 [ ] Doesn't speak another language_
C9.	What is your highest education level?	0. [ ] No formal education 1. [ ] Primary (1-8 grade) |__|__| 2. [ ] Secondary (9-12 grades): |__|__| 3. [ ] Diploma or Certificate 4. [ ] University 5. [ ] Post-graduate studies 6. [ ] Informal education
C10.	Can you write a letter in the language you speak at home?	1. [ ] Yes 0. [ ] No
C11.	Can you read a letter in the language you speak at home?	1. [ ] Yes 0. [ ] No
C12.	Which local word do you use to express the measurement of your farm land? (FO: this is a very important question!!! - It will be preloaded in the next questions and sections)	1. [ ] Kada 2. [ ] Timad 3. [ ] Gasha 4. [ ] Gezim 5. [ ] Gemed 99. [ ] Other (specify)_____
C13.	How many l**${(Local_C12)}** make an hectare?	_____ [ ] -98 Don't Know
C14.	If the DA recommends you to apply 100 kg of fertilizer per hectare, how many kilograms you should apply in one **${(Local_C12)}**?	_____ [ ] -98 Don't Know
C15.	What is your relationship to the household head? You would say that you are (READ OPTIONS) of the household head?	0. [ ] I am the household head 1. [ ] Spouse 2. [ ] Daughter / Son 3. [ ] Brother / Sister 4. [ ] Niece / Nephew 5. [ ] Grandchild 99. [ ] Other (specify):_____
C16.	Are you the person who buys most of the agricultural inputs for your plots? *[FO: Explain that this question is asking about the person who goes to the shop to purchase seeds, fertilizers, pesticides, and/or herbicides]*	0. [ ] No 1. [ ] Yes 2. [ ] Yes, jointly with my spouse 3. [ ] Yes, jointly with my parents/elders in the household

C17.	When you want to call someone, do you dial the number yourself or you ask someone else to do it for you?	1. [ ] Yourself -> **C19** 2. [ ] Someone else dials for you
C18.	Who helps you to dial when you want to make a phone call?	1. [ ] Spouse 2. [ ] Daughter / Son 3. [ ] Brother / Sister 4. [ ] Nice / Nephew 5. [ ] Grandchild 99 [ ] Other (specify):_____
C19.	Are you the person who normally recharge the balance of this phone by either dialing *805 or 909?	1. [ ] Yes 0. [ ] No

SECTION D: LR 2017 FARMING PRACTICES

[FO: Read] "Now I am going to ask you some questions about your plots and your farming practices in the Rain Season that has just finished, ***last Rain Season.****"*		
D1.	How many **${(Local_C12)}** of land do the members of your household and you planted, either owned or rented)? *[FO: Clarify to the respondent to report the total size of ALL the land farmed by the household.]*	_____ LOCAL MEASUREMENT → go to **Section E if answer 0**
D2.	Now, I am going to ask you about your cereal crops. Which cereals did your household plant **this last Rain Season**? *[FO: first mark the cereal crops that were planted, and then ask the total land planted for each cereal, the total number of bags harvested and the number of kilograms of those bags]*	
	1. [ ] Barley D2.1.1 Total land area ____C16 UNIT D2.1.2 Did you finish harvest? [ ] 1. Yes [ ] 0. No D2.1.3 Total # of bags harvested (or expect to harvest harvest) ___ D2.1.3 How many kilograms per bag___ 2. [ ] Maize D2.2.1 Total land area ____C16 UNIT D2.2.2 Did you finish harvest? [ ] 1. Yes [ ] 0. No D2.2.3 Total # of bags harvested (or expect to harvest harvest) ___ D2.2.4 How many kilograms per bag___ 3. [ ] Sorghum D2.3.1 Total land area ____C16 UNIT D2.3.2 Did you finish harvest? [ ] 1. Yes [ ] 0. No	4. [ ] Tef D2.4.1 Total land area ____C16 UNIT D2.4.2 Did you finish harvest? [ ] 1. Yes [ ] 0. No D2.4.3 Total # of bags harvested (or expect to harvest harvest) ___ D2.4.4 How many kilograms per bag___ 5. [ ] Wheat D2.5.1 Total land area ____C16 UNIT D2.5.2 Did you finish harvest? [ ] 1. Yes [ ] 0. No D2.5.3 Total # of bags harvested (or expect to harvest harvest) ___ D2.5.4 How many kilograms per bag___ 6. [ ] Millet D2.6.1 Total land area ____C16 UNIT D2.6.2 Did you finish harvest? [ ] 1. Yes [ ] 0. No

	D2.3.3 Total # of bags harvested (or expect to harvest harvest) ___ D2.3.4 How many kilograms per bag___	D2.6.3 Total # of bags harvested (or expect to harvest harvest) ___ D2.6.4 How many kilograms per bag___ 7. [ ] Sesame D2.7.1 Total land area ____C16 UNIT D2.7.2 Did you finish harvest? [ ] 1. Yes [ ] 0. No D2.7.3 Total # of bags harvested (or expect to harvest harvest) ___ D2.7.4 How many kilograms per bag___
	FO read: "Now I am going ask you questions about your practices of the plot that was planted with **${(NAME OF CHOSEN CROP)}.** FO: FOR THE NEXT SET OF QUESTIONS THE TABLET WILL CHOOSE THE CROP THAT OCCUPY THE LARGEST AMOUNT OF LAND (ie **${(NAME OF CHOSEN CROP)}**. IF THERE IS A TIE, THE TABLET WILL RANDOMLY PICK ONE OF THE CROPS.	
D3.	What did you do with the crop residues from the previous season?	1. [ ] incorporated to the land 2. [ ] feed the animals 3 [ ] use for fire 4 [ ] sell them
D4.	How many times your land was ploughed before planting **${(NAME OF CHOSEN CROP)}** during the last Long Rain Season?	_____ times
D5.	Were your ${**(NAME OF CHOSEN CROP)}** plot planted manually, using oxen, or machinery?	1. [ ] Manually 2. [ ] Oxen 3. [ ] Machinery
D6.	Were organic fertilizer/compost applied in the land where you planted ${**(NAME OF CHOSEN CROP)}** last Long Rain Season?	1. [ ] Yes 0. [ ] No
D7.	Were modern chemical fertilizers applied on your **(NAME OF CHOSEN CROP)** plot?	1. [ ] Yes 0. [ ] No **-> D11**
D8.	How many times did you apply modern chemical fertilizers your ${**(NAME OF CHOSEN CROP)}** plot?	_____ times
D9.	Did you apply modern chemical fertilizers at planting in your ${**(NAME OF CHOSEN CROP)}** plot?	1. [ ] Yes 0. [ ] No
D10.	Which chemical fertilizers were applied in your${**(NAME OF CHOSEN CROP)}** plot, how many kilograms of each one in total? *[FO: Listen and mark all that apply and complete the amount of kilograms only for the fertilizers that were applied]*	1. [ ] DAP (black) (**D10.1.1)** Total Amount__kg 2. [ ] NPS (grey) / blended fertilizer (**D10.2.1)** Total Amount_____kg 3. [ ] UREA (white) (**D10.3.1)** Total Amount_____kg 4. [ ] Other 1 (specify) (**D10.4.1)** Total Amount_____kg 5 [ ] Other 2 (specify) (**D10.5.1)** Total Amount_____kg
D11.	Did you use your own ${**(NAME OF CHOSEN CROP)}** seeds or did you buy them at the Cooperative Store, or Private Dealers?	1. [ ] Own seeds 2. [ ] Cooperative Store 3. [ ] Private Dealer/Local market 4. [ ] Borrowed seeds

	[FO: mark all that apply.]		
D12.	Were the seeds planted in rows or by traditional broadcasting in the ${**(NAME OF CHOSEN CROP)}** plot?	1. [ ] rows 2. [ ] traditional broadcast	
D13.	Were any chemical pesticides on your ${**(NAME OF CHOSEN CROP)}** plot?	1. [ ] Yes 0. [ ] No	
D14.	How many times did you weed your ${**(NAME OF CHOSEN CROP)}** plot last season?	_____times	
D15.	Were herbicides applied to your ${**(NAME OF CHOSEN CROP)}** plots last season?	1. [ ] Yes 0. [ ] No	
D16.	Did you and/or are you planning to apply any chemical pesticides after harvest on your ${**(NAME OF CHOSEN CROP)}** on produce?	1. [ ] Yes 0. [ ] No	
D17.	Which others crops did your household plant this last Rain Season 2017? *[FO: listen and mark all that apply]*	**Pulses, Oil Seeds and Cash Crops** 1. [ ] Sesame 2. [ ] Fava Bean 3. [ ] Chickpea 4. [ ] Common Bean 5. [ ] Cotton 6 [ ] Rice 7 [ ] Lentils	**Vegetables** 8. [ ] Potato 9. [ ] Carrot 10. [ ] Onion 11. [ ] Garlic 12. [ ] Tomato 13. [ ] Pepper 14. [ ] Head Cabbage **Fruits** 15. [ ] Bananas 16. [ ] Mango 17. [ ] Avocado 99 [ ] Other -97 [ ] No other crop was planted
D18.	In general, what are the three most regular problems you faced with your crops that reduce your yield? Please name them by order of importance *[FO: choose three and specify the rank*	1. [ ] Rain shortage___rank 2. [ ] Floods___rank 3. [ ] Lack of soil fertility 4. [ ] Worms/Pests __ rank 5. [ ] Weeds ___ rank 6. [ ] Shortage of seeds at the store__rank 7. [ ] Shortage of fertilizers at the store_rank 8. [ ] Shortage of pesticides/herbicides at the store_rank 9. [ ] Lack of credit to buy inputs 10. [ ] Lack of Machinery / Oxen	
D19.	In order to solve these problems, who are your three primary source of information? *[FO: choose three and specify the ranking)*	1.[ ] Woreda Expert _____rank 2.[ ] DA _____rank 3.[ ] Model Farmer _____rank 4.[ ] Other farmers _____rank 5 [ ] Family members _____rank 6 [ ] Radio _____rank 7 [ ] TV _____rank 8 [ ] 8028 _____rank 9 [ ] Social Media _____rank 10 [ ] Other (specify) _____rank	

SECTION E: EXPERIENCE WITH 8028 HOTLINE

FO: Now, I would like to ask you some questions about the the 8028 hotline, that offers agricultural recommendations to farmers in your area

Screening

E1.	Have you ever heard about the 8028 hotline?	1. [ ] Yes 0. [ ] No **-> E4**
E2.	How did you hear about the 8028 Hotline **for the first time**? *[FO: listen and mark all that apply]*	1. [ ] Radio 2 [ ] Development Agent 3 [ ] Fellow farmer 4 [ ] Adult relatives 5 [ ] Children 6 [ ] Registration Bookmark/Pamphlet 7 [ ] Other (specify):_____
E3.	Do you think you have to spend your phone credit to call to the 8028 hotline?	1. [ ] Yes 0. [ ] No
E4.	Have you or someone used this phone to call into the 8028 hotline to access recommendations on agricultural practices?	1. [ ] Yes 0. [ ] No **FO: if the respondent answered NO to E1 and E4 -> Section F Alerts**
E5.	Who accessed the 8028 hotline?	0. [ ] Myself -**> E9** 1. [ ] Spouse 2. [ ] Daughter / Son 3. [ ] Brother / Sister 4. [ ] Nice / Nephew 5. [ ] Grandchild 99[ ] Other (specify):_____
E6.	Have you listened to the recommendations with this person?	1. [ ] Yes -**> E9** 0. [ ] No
E7.	When could I talk to this person?	1 [ ] Now 2 [ ] Later (specify **E7.B)** 1 [ ] Date (calendar): 2 [ ] Time: 1. [ ] Between 8am – 11am 2. [ ] Between 11am – 2pm 3. [ ] Between 2pm – 5pm **-> FO: continue with this respondent on Section F Alerts**
E8.	**FO: Introduce yourself and the study to the new respondent** My name is _____I am calling on behalf of the Agriculture Transformation Agency (ATA). Currently, we are speaking to different users of the free 8028 hotline that offers access to relevant information to improve their farming practices, The user of this phone told me that you have called the 8028 hotline before and we would like to know more about your experience with	1. [ ] Now 2. [ ] Later (specify **E8.1)** 1. [ ] Date (calendar): 2. [ ] Time: 1. [ ] Between 8am – 11am 2. [ ] Between 11am – 2pm 3. [ ] Between 2pm – 5pm **-> FO: continue with the first respondent on Section F Alerts** **-**

	the system. Do you have 5 minutes to answer some questions?	

Question about usage

E9.	Has someone ever taught you to use the 8028 hotline?	1. [ ] Yes 0. [ ] No
E10.	Who taught you?	1. [ ] Trained by DA / Woreda Expert 2. [ ] Trained by other farmer 3. [ ] Trained by teacher 4. [ ] Trained by children 99. [ ] Other
E11.	Why do you normally call the system? *[FO: listen and choose all that apply]*	1. [ ] To learn about modern farming practices 2. [ ] To learn about new crops 3. [ ] To learn about weather 99. [ ] Other
E12.	Have you discussed the 8028 recommendations with someone? If so, who did you discuss?	1. [ ] Yes -> E12.1 Who? 1. [ ] DA 2. [ ] Fellow Farmer 3. [ ] Household Member 4 [ ] 5 to 1 group 5 [ ] Other 0. [ ] No
E13.	Who do you think put these recommendations together?	1. [ ] The Agricultural Transformation Agency (ATA) -> E13.1 Do you think that the Agricultural Transformation Agency is party of the Government or an NGO? 1. [ ] The Ethiopian Government 2. [ ] An NGO 2. [ ] The Ethiopian Government 3. [ ] Ministry of Agriculture 4. [ ] Universities / Research Centers / Researchers 5. [ ] An NGO 6. [ ] Ethiotel 99. [ ] Other (specify): _____ [ ] -98 Don't Know
E14.	How much do you trust recommendations from 8028 hotline? [FO: PLEASE READ THE OPTION[	1. [ ] Never trust the recommendations at all 2 [ ] Hardly trust t the recommendations 3 [ ] Mostly trust the recommendations 4 [ ] Always/Completely trust the recommendations

E15.	What are the major barriers you face to follow the recommendations from the 8028 hotline? [FO: LISTEN AND MARK ALL THAT APPLY)[	1. [ ] Couldn't adapt the units of measurement to the land size 2. [ ] Didn't understand with the recommendation 3. [ ] Didn't agree with the recommendation 4. [ ] Didn't know where to buy the products they recommended 5 [ ] Didn't have the money 6. [ ] Prices of the inputs were too high . 99. [ ] Other ,specify (**E15.A**)_____ -97 [ ] Didn't face major barriers
E16.	Now, I will ask you to think about the famers in your area. Out of 10 typical farmers in your village, how many do you think are using the system very well?	_____ out of 10
E17.	What are the three main issues/limitations farmers in your area facing to use the system as well as you related to accessing the system, using their phone, literacy, language, listening to the recommendations, etc.)	**Problems accessing the system and/or us phones** 1. [ ] Network connectivity 2. [ ] No reliable source of electricity 3. [ ] Do not have working phones 4. [ ] No credit 5. [ ]/buttons not working **Literacy problems** 1. [ ] Low level of numerical literacy, cannot recognize numbers 2. [ ] Don't speak the languages available in the system 3. [ ] Lack of ability to use a phone 4. [ ] IVR concept unfamiliar, hard to learn without training **Problems with the recordings/menus** 1. [ ] Can't complete profile question 2. [ ] Recommendations/Menu recordings spoken too fast, 3. [ ] Mistakenly chose a different language and can't undo 4. [ ] Menus where two digits have to be pressed (e.g. "07" for some woreda, or crop, or region) 5. [ ] Farmers don't know what * and # mean **Problems with the content of the recommendations** 1. [ ] Some terms in the recommendations unfamiliar 2. [ ] Some menu options have unclear or multiple meanings/interpretations 3. [ ] Unfamiliar measurement units used 4. [ ] Too many detailed mentioned in content messages (too many numbers, product names not always familiar) 5. [ ] Difficulty adapting measurement units (given per hectare) to smaller land sizes **Problems implementing the recommendations**

		1. [ ] Lack of money / finance to adapt the recommendations 2. [ ] Recommended inputs unavailable locally
E18.	Have you ever tried to train other farmers to use the system?	1. [ ] Yes 0. [ ] No

SECTION F: EXPERIENCE ALERT SYSTEM

FO: Now, I would like to ask you some questions about the 8028 alert		
F1.	Have you received a call from the 8028 system with a voice recorded message? If so, did you answer the call?	1. [ ] Yes, but didn't pick up 2. [ ] Yes, I answer the call → **F3** 0. [ ] No → **F7** -98. [ ] Don't know/ Don't remember → **F7**
F2.	Why didn't you answer the call? *[FO: listen and mark all that apply]* **Answer the question and skip to → F7**	1. [ ] I thought that it was a scam call 2. [ ] I thought that it was Ethiotel 3. [ ] I thought they will charge me 4. [ ] Thought it was from abroad. 5. [ ] Other:_____ -98. [ ] Don't know
F3.	What was the voice recorded message about? *[FO: listen and mark all that apply]*	1. [ ] Drought 2. [ ] Flood 3. [ ] Disease outbreak 4. [ ] Pest outbreak -98. [ ] Don't know
F4.	Did you share the message of the alert with anybody?	1. [ ] Yes 0. [ ] No → **F7** -98. [ ] Don't know/Don't remember → **F7**
F5.	Who did you talk with about the content of the alert? *[FO: Don't read question options and mark all that apply.]*	1. [ ] Other farmers 2. [ ] Development Agent 3. [ ] Neighbors 4. [ ] Friends 5. [ ] Family members 99.[ ] Other, specify: _____
F6.	How clear are the recommendations received?	1. [ ] Very unclear 2. [ ] Unclear 3. [ ] Clear 4. [ ] Very clear
F7.	Have any of your friends, relatives or neighbor received a call from the 8028 hotline with recorded message that provide an alert of an upcoming weather, disease and/or pest alert?	1. [ ] Yes 0. [ ] No → **End** -98. [ ] Don't know/Don't remember → **End**
F8.	Did they share that information with you?	1. [ ] Yes 0. [ ] No -98. [ ] Don't know/Don't remember

SECTION Z: END

[FO: Read] "*Thank you for agreeing to participate in this short survey."*		
Z1.	Should this person be called back?	1. [ ] Yes → **Z3** 0. [ ] No
Z2.	Why should this person not be called back?	
Z3.	FO: Please record any comments or notes about this survey.	
Z4.	End Time	|___|___| : |___|___|

C Declaration of generative AI and AI-assisted technologies

During the preparation of this work the authors used ChatGPT in order to edit the final text of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.