FEED THE FUTURE

The U.S. Government's Global Hunger & Food Security Initiative

EVALUATION OF THE IMPACT OF E-VERIFICATION ON COUNTERFEIT AGRICULTURAL INPUTS AND TECHNOLOGY ADOPTION IN UGANDA

Endline Report

August 2019

Daniel O. Gilligan, Naureen Karachiwalla, and Giang Thai

USAID FROM THE AMERICAN PEOPLE

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

IFPRI® Supported by the CGIAR

An Evaluation of the Impact of E-verification on Counterfeit Agricultural Inputs and Technology Adoption in Uganda

Endline Report

Daniel O. Gilligan

Naureen Karachiwalla

Giang Thai

August 15, 2019

Delivered to the United States Agency for International Development
for the Feed the Future Initiative

International Food Policy Research Institute
1201 I St. NW
Washington, DC 20005
USA

Executive Summary

Introduction

Adoption of productivity- and income-enhancing agricultural inputs is low, particularly in Africa, despite the importance of farm incomes in rural welfare (Aker, 2011; Duflo et al., 2011; Foster and Rosenzweig, 2010). In Uganda, use of high-quality agricultural inputs such as hybrid maize seed, glyphosate herbicide, and inorganic fertilizer is extremely low (Ashour et al., 2015). This leads to low agricultural productivity and limits incomes. Low levels of technology adoption are compounded (and may partially be explained by) a lack of farmer trust in the current inputs supply system, which has been plagued by counterfeiting (Fairbairn et al., 2018; Bold et al., 2018).

Counterfeit products range from benign fake or adulterated materials, to banned substances that are harmful to both crops and human health. Counterfeiters have become increasingly innovative in their techniques, making it difficult to identify their products without laboratory tests, while farmers and agro-dealers have little means of verifying whether a product is genuine, unexpired, priced fairly, or accurately labeled by brand, type, or concentration.

To help mitigate this issue, USAID through the Feed the Future (FTF) initiative has supported a government program for input authentication called e-verification (EV). E-verification involves labeling agricultural inputs with a scratch-off label that provides an authentication code that can be used to confirm that the labeled product is what is claimed on the packaging (for example, brand, bottle size). The consumer sends the code that they have scratched off the product to a short code and receives back an SMS message confirming the identity of the product.

The objectives of the e-verification project are to increase adoption of productivity-enhancing agricultural inputs, to increase farmer productivity and profits as a result of adoption, and ultimately, to reduce the prevalence of counterfeit and adulterated agricultural inputs on the market. The impact evaluation estimates the impact of the e-verification scheme on each of these outcomes and examines how the e-verification project achieved its results. The three main agricultural inputs that we study are hybrid maize seed, glyphosate herbicide, and inorganic fertilizer.

As stated in the Baseline Report, the primary research questions that this impact evaluation will address are as follows:

1. Does improving access to e-verified agricultural inputs result in greater take-up of high-quality inputs by farmers?
2. What is the impact of increased use of agricultural inputs on incomes and yield for Ugandan farmers?
3. Did the e-verification program contribute to a reduction in counterfeiting and adulteration of agricultural inputs in the market?

This endline report presents the results of the study. It provides the details of the study design, interventions, sampling, and estimates of the causal impact of the e-verification program on farmer outcomes.

The E-verification sub-activity

The objective of the EV initiative in the FTF Ag Inputs activity is to increase production of maize, beans, and coffee through the appropriate use of high-quality agro-inputs (seeds, fertilizers, herbicides, and pesticides).

The impact evaluation will focus on ascertaining the impact of two of the four Intermediate Results (IR 1 and 2) of USAID/Uganda's Agriculture Inputs Activity:

- IR1 "Increased availability of high-quality inputs to farmers in FTF focus districts"
- IR2 "Decreased prevalence of counterfeit agricultural inputs"

In terms of FTF objectives, the project relates to those of improving agricultural productivity and increasing private investment in agriculture. The impact evaluation tests the following hypothesis: "If USAID/Uganda introduces and promotes electronic verification (EV) and effective marketing of agricultural inputs, then Ugandan farmers in key FTF districts will demonstrate higher adoption rates of those inputs, resulting in a commensurate improvement in agricultural productivity."

The Uganda National Bureau of Standards (UNBS), which has the mandate to ensure quality standards are met nationally across a wide array of products, implemented the e-verification scheme under the name 'E-tag', also known as 'Kakasa'. Later, certain seed companies also implemented a slightly different tag under the E-tag scheme, called Ag-verify. A third scheme called CHK tag was provided for one of the glyphosate herbicide companies as well, but it did not last very long, as the E-tag system had more success.

Encouragement Design and Interventions

The theory of change is as follows. If farmers suspect that inputs may be fake or adulterated, this causes them to reduce their demand for the products. Farmers who would otherwise use the products may avoid them altogether. Reduced demand arises because the uncertainty about the products' authenticity reduces farmers' expected returns from adopting the products. This lack of demand reduces input prices and harms the profitability of selling genuine high-quality inputs. As a result, many such genuine inputs may cease to be available in the market, leaving only inputs of suspect quality or low quality. This problem of missing information about a product harming demand is known in economics as "adverse selection." The counterfeiting problem is a form of the adverse selection problem described by Akerlof (1970) in his paper on the effect of unobserved quality of used cars, "The Market for `Lemons'." If the quality of remaining products on the market becomes known to farmers through experience, prices may adjust to reflect the lower returns to using these counterfeit products (some versions provide some productivity boost), but this still reduces availability of higher quality, productive technology that could raise incomes. The EV sub-activity could help to overcome this problem by providing a form of input authentication with information for agricultural inputs. This would help to increase demand for those inputs. Improved take-up of genuine productivity-enhancing inputs should help in increasing yields, farm productivity, and household income.

The impact evaluation of the EV component to the Ag Inputs activity uses an encouragement design to causally identify the impacts of access to e-verified inputs on take-up of high-quality inputs, prevalence of counterfeiting, and farmers' yield and net income for maize. Markets gain access to agricultural products through a variety of conduits, including directly from suppliers, through distributors reaching out to retail markets, and by retailers visiting distributors based in major cities. This makes it impossible to systemically assign retail markets into treatment and control groups, thus ruling out a randomized control trial (RCT) design as a method for identifying impacts.

Several studies in the economics literature have used encouragement designs to identify the impacts of interventions that would not otherwise be possible using RCTs (Duflo and Saez, 2003; Duflo, Kremer, and Robinson, 2011; Thornton, 2008; Katz, Kling, and Liebman, 2001). Studying adoption (be it agricultural inputs or other programs) is particularly well-suited to this approach. When products or programs are available to everyone, random variation in exposure can be induced by encouraging some people to adopt and not others.

By randomly assigning one of two matched communities in each market location to receive 'encouragements' – in our case interactive voice recording (IVR) messages and community promotion meetings (CPMs) – the encouragement design produces exogenous variation in awareness about the availability of e-verified products. This difference in awareness should produce higher take-up of e-verified products in encouragement communities than in non-encouragement communities. Thus, the randomized encouragement campaign produces an experimentally induced wedge in awareness of e-verified products in encouragement treatment communities which can be used to explain the effect of e-verification on take-up of these inputs and other farm outcomes. The experimental design makes it possible to conclude that any difference in take-up of hybrid maize seed or glyphosate herbicide at endline, for example, between encouragement and non-encouragement communities must be due to differences in exposure to e-verification. Without that randomized exposure, we should expect use of quality inputs to be the same on average between encouragement and non-encouragement communities.

As part of the encouragement campaign, two rounds of IVR messaging and CPMs were conducted: the first was in the second growing season of 2016 and the second was in the first growing season of 2017. The IVR messages explained what Kakasa products are, why they exist, and how to use them, and included a reminder always to authenticate. The message then gave the listener the option to press 1 to hear a list of Kakasa herbicide that would be available, and to press 2 to hear a list of Kakasa maize that would be available.

The Community Promotion Meetings were held in a centrally located place within the community. Upbeat music was played while people arrived, and raffle tickets were handed out on arrival. The village leader provided an introduction to the team, and a skit was performed by the pair of enumerators. The enumerators then conducted a demonstration of Kakasa products. There was a question-and-answer session where correct answers would win people prizes, and there was a raffle to give away Kakasa products.

During the second round of CPMs in the first season of 2017, vouchers were handed out at the end of the meeting. Meeting participants were told that each person attending the meeting would receive two vouchers: one for purchasing Kakasa herbicide and one for purchasing Kakasa maize. Community members were informed that the two enumerators would return to the community in approximately three weeks to provide redemptions for the vouchers. In order to redeem a voucher, the individual needed to provide the original voucher, and proof of purchase (this could be a receipt, an empty container, or a full container). If the tag was not scratched to authenticate, the enumerator would do so with them upon redemption of the voucher. Enumerators kept track of how many vouchers were given out in each community and how many were redeemed.

Sample Design

The first step of the sampling process was identifying broad areas in which to work. IFPRI worked with Tetra Tech to identify "market hubs" where e-verified products were likely to be made available. A market hub is a major market area consisting of a collection of "market locations" covering one or more districts, centered around a major town (generally, the district town center). A market location is a collection of retail shops (often referred to as a trading center) selling agricultural inputs in the same vicinity. Each market location serves several surrounding villages. Ten market hubs were identified alongside Tetra Tech for inclusion in a market survey. These are: Hoima, Iganga, Kasese, Kiboga, Luwero, Masaka, Masindi, Mbale, Mityana, and Mubende. They were selected according to two main criteria: location in a high maize growing area, and probability of receiving e-verified products.

A market survey was conducted that gathered lists of the market locations served by these market hubs, of villages served by each market location, and of retail shops in each market location. Using the data gathered during the market survey, 120 market locations were selected for inclusion in the study. The market survey also collected data on each village surrounding the market hub, including village population (number of households), proportion of farmers growing maize, and distance to the village from the center of the market hub. From this information, villages were matched into pairs based on degree of similarity in these three criteria, and this yielded a sample of 240 villages.

Randomized assignment was used to allocate one village in each pair to the encouragement treatment. The other village in each pair serves as the control village, from the standpoint of the encouragement design. Both treatment and control villages would have access to the e-verified products available in their local market location or from any other market products. Further, of the 120 encouragement villages, 60 villages were randomly selected to receive price discounts for purchasing EV maize and herbicide. In treatment villages, all households that had a phone received the encouragement treatment in the form of a series of interactive voice recordings (IVR) informing them about the availability of e-verified products in their local market. Encouragement villages were also visited by the survey team who would conduct Community Promotion Meetings (CPM), whereby they would advertise Kakasa products and conduct demonstrations of how to use them. Control villages would have *access* to the e-verified products, but no encouragement messages would be sent to households in those villages, no meetings would be held in these villages, and no discounts would be offered.

For each of the 240 LC1s that were selected for inclusion into the study, a community listing exercise (CLE) was conducted, whereby enumerators attempted to visit all households residing in the LC1. From the community listing exercise, ten households were randomly selected to receive an extensive household survey collecting information on many topics. Seven households who reported in the CLE that they owned a mobile phone were randomly selected for inclusion into the household sample, and three households who reported in the CLE that they did not own a mobile phone were also randomly selected for inclusion into the household sample. These households were sampled using simple random sampling.

Surveys

Three rounds of data collection were undertaken: baseline, baseline II, and endline. A second baseline was conducted to update key variables since the implementation of EV was delayed. The baseline survey was conducted in July-August 2014, the second baseline in January 2016, and the endline in July-August 2017. The endline was conducted two 'growing seasons' after the intervention began and after two seasons of implementation of the encouragement activities.

The baseline data were collected using a series of four survey instruments to gather data at the market level, the village level, and the household level: (1) the market survey was used to identify market locations in the study market hubs, collect shop-level data from all retail shops selling agricultural inputs in each market location, and identify the villages served by each market location; (2) the Community Listing Exercise (CLE), which briefly interviewed all households in the LC1s that were randomly selected for the baseline survey; (3) the household survey was designed to collect information at the household, individual, and plot levels from a sample of ten households in each study village; and (4) the community survey collected information on demographics, land and infrastructure from the LC1 chairperson in each study community.

The shop questionnaire consisted of general characteristics of the shop (type of building, etc.), products and sales, demographic information on the shop owner, banking and credit, products sold, their source, their price, and the top brands of maize, herbicide, and fertilizer, and farmer perceptions and recommendations made to farmers. During the CLE, data were collected on household demographics, farming and input use, purchases of agricultural inputs, and access to and use of mobile phones. Data were collected on household demographics, farming and input use, purchases of agricultural inputs, and access to and use of mobile phones.

Ten farmers in each village also received the household survey instrument. The complete list of modules included in the household survey instrument is provided in the baseline report, but data collected included: a household roster to collect basic demographic information on all current household members such as age, marital status, relationship to household head, and education. Information on agricultural production was asked regarding each parcel farmed by the household. Of the plots listed with any maize production during the current agricultural season, the program randomly selected one plot as a representative maize plot (RMP). The respondent was asked about production details and sales of all crops from the RMP. The production module also collected data on input use and perceptions of input quality for all plots during the current and previous agricultural seasons and specific input use on the RMP for the current season. Other modules used to assess household wealth, assets, and vulnerability were also included, as was a food consumption module.

Data on individual preferences (risk and ambiguity aversion) and beliefs regarding counterfeiting were also collected. These modules included both qualitative and quantitative questions asked of the primary agricultural decision maker (PADM) and the secondary agricultural decision maker (SADM). The qualitative questions were administered during the main part of the interview. The quantitative data was collected after the conclusion of the main survey in a group format with all available PADM respondents from the village (maximum 10 respondents). Each game presented a hypothetical agricultural scenario and respondents were asked to make a choice that would identify their risk or ambiguity preferences. Respondents were also asked about their beliefs of the quality of different agricultural inputs on the market using props to form a distribution to represent their confidence in the quality of the products. For further details on the post-interview games, see Ashour et al. (2018).

The community questionnaire was administered at the time of the household questionnaire by the team leaders who interviewed the LC1 chairperson of each village. Topics covered in the community questionnaire included demographics, land, infrastructure, farming activities, and major events experienced by the community, both positive and negative, in the past two years.

The second baseline updated information on certain key variables that were needed, both to understand adoption patterns as well as to update phone numbers since this was the primary means of providing the encouragement intervention. All households were asked demographic questions about the household head and PADM, basic questions on farming, perceptions of input quality and counterfeiting, risk and ambiguity aversion, and phone numbers. The ten household survey households were additionally asked questions about input use and purchase of hybrid maize, glyphosate herbicide, and inorganic fertilizer.

A short shop questionnaire was also administered to all shops that existed at that time in each of the 120 market locations. This was a shortened version of the baseline shop questionnaire, and also checked whether any shops had closed down and if any new shops had opened.

The endline survey followed a similar structure to the baseline survey. A CLE survey was administered to all households that were listed as part of the first two CLE surveys, a household survey was administered to all households who were listed as household survey households in the baseline survey or second CLE, and a community survey was also conducted.

The endline CLE and household surveys included all of the questions from the second CLE and the baseline household survey and also included a detailed section on experience with EV products and with the encouragement interventions. The short shop questionnaire from the second round of data collection was also administered in the shops currently in operation in all 120 market locations.

Estimating Treatment Effects

Individual and household characteristics, coverage of other interventions, and community characteristics should be equal, on average, across both the encouragement and control treatment arms and within and between the discount and no discount arms and the discount arms compared to the control group as a result of randomization. Consequently, average differences in outcomes across the treatment arms after intervention can be interpreted as causal impacts; that is, they can be interpreted as being caused by, rather than simply correlated with, the interventions.

The main outcomes for this report include:

1. The rate of take-up of improved agricultural inputs (glyphosate herbicide, inorganic fertilizer, and hybrid maize seeds);
2. The level of farm yields, incomes, and profits; and
3. Measures of input quality for hybrid maize seed and glyphosate herbicide.

We estimate two models to measure the difference in average outcomes between encouragement and control communities. The first is a "reduced form" model, which estimates the *impact of the encouragement activities* on study outcomes (the direct effect of the interactive voice messages and community meetings provided through the encouragement campaign). This model reports the average effect of being assigned to an encouragement village (and thus being exposed to the encouragement interventions) on adoption of high-quality inputs (hybrid maize, herbicide, and fertilizer) as well as on farm level outcomes such as yields, farm income, and net income.[1] It tells us whether the encouragement activities on their own had an impact on adoption and farm outcomes. We estimate two versions of this reduced form model: (i) without including household level baseline control variables (reported in the main impact estimates in the body of the report), and (ii) including household level baseline control variables (reported as a robustness check, in Appendix A).

The second model uses an "instrumental variables" approach to estimate the causal impact of e-verification on adoption of high-quality inputs and other farm outcomes, which addresses the main policy question for this report. More specifically, the model estimates the impact on input use and farm outcomes of being aware of e-verified products, based on experimental variation in this awareness created by the encouragement campaign. This tells us the *impact of awareness of the e-verification product assurance scheme itself* on adoption and farm outcomes. This approach provides "Local Average Treatment Effect (LATE)" estimates, which means it estimates the impact of the e-verification intervention on adoption and farm outcomes for those households in encouragement communities who are induced to take-up the product due to the encouragement intervention (not for those who would have taken it up anyway or for those who would never take it up). We do this because people in the control group also had access to EV products and some adopted them.

This model is estimated in two stages (two regressions). First, we estimate the effect of the encouragement interventions on awareness of e-verified inputs and obtain an estimate of the predicted probability that a household is aware of e-verified products. The encouragement interventions are implemented to cause higher awareness in encouragement communities than control communities in a randomly assigned way. In the second stage, we estimate the effect of the predicted probability of being aware of e-verified products from the first stage regression on outcomes such as farm incomes.

This two-stage model of the randomized encouragement design is needed to obtain unbiased estimates of the impact of e-verification because, without the encouragement campaign, awareness of e-verified products is not random. Households that are aware of e-verified products or have tried them on their own are likely to have different characteristics than households on average, such as being more innovative or having better knowledge of new technologies. Comparing outcomes for these "innovative" households to those who are not aware of e-verified products or have not tried them would lead to biased impact estimates as the difference in outcomes would include the effect of households using e-verification being more innovative. A common approach to designing a study to remove the effect of these characteristics is to conduct a randomized control trial, where

[1] The reduced form model is an "intent-to-treat" model of the impact of the encouragement campaign, as it measures the average effect of being exposed to the randomly assigned encouragement campaign. This model does not provide intent-to-treat estimates of the impact of the e-verification scheme itself.

e-verified products are made available only in randomly selected markets. However, it was not possible to control which markets got access to e-verified products (all households in the study did). The randomized encouragement design is a good alternative study design in this case because the randomly assigned encouragement campaign creates experimental variation in awareness or exposure to the e-verified products. This should lead to higher take-up of these products in randomly assigned encouragement communities than in randomly-assigned control communities around the same markets. We expect to observe awareness or adoption of e-verified products in both encouragement treatment and control communities, but higher adoption in encouragement treatment communities. In the model, the random variation comes from the first stage equation – the encouragement to be more aware. If we estimate the second stage without estimating the first stage, our impact estimates would be biased.

At baseline, the goal was to interview 2,400 households on the complete baseline household survey. We were able to successfully interview 2,375 households (some households on the sample lists refused or were not available and could not be replaced due to time constraints). We have data on 2,047 households at endline. This implies a rate of attrition of 14 percent.

We find that there is no differential attrition by treatment group between the first and second baseline rounds. Between the baseline and endline rounds, households in the encouragement (but not discount) treatment group were 2 percentage points less likely to be in the sample at endline. Although this result is statistically significant, it is very small.

Results

Baseline Balance

At baseline, we find that households across encouragement treatment arms are very similar to one another. We have very good balance on household demographic characteristics. Slightly more households in the control group have female primary agricultural decision makers, and the household durables asset index is lower in the control group compared to the other groups (however the sub-indices are well balanced), and this then drives a slight imbalance in the overall asset index as well. However, when conducting 42 tests, we do expect that approximately 2 of them will be significantly different by chance.

With regards to farming, on average, households own two parcels of land and own approximately 5-6 acres of land. Households grow an average of five crops during the first season of 2014 and about four different crops during second season of 2013. Use of hybrid maize seed is relatively low with only about ten and eight percent of the sample using hybrid maize in first season 2014 and second season 2013, respectively. However, we do see that almost all households in the sample (99 percent) grew maize in first season 2014 and approximately 93 percent of the sample grew maize in the second season of 2013. Similar patterns are observed in the rest of the variables as well. Once again, our baseline characteristics regarding farming and input use are fairly well balanced. There are slight differences between groups in terms of use of fertilizer in first season 2014 and use of hybrid and any maize seeds in second season 2013, however, again, with 48 tests involved what we find is approximately what we should find statistically by chance.

Experience with EV and the Encouragement

Very few people in the control group recognize the E-tag or Ag-verify labels, and only 2 percent had heard about E-tag. In the encouragement and discount groups, however, almost half of households recognized the label and two thirds had heard of E-tag. Most people had heard about E-tag through community meetings (these include the community promotions we had organized as part of the encouragement; AgInputs and E-tag also had road shows) and most people heard about Ag-verify through phone calls. Most people first heard about E-tag in the first season of 2016 when the products were first launched.

When we examine purchasing decisions we see that about 20 percent of households in the encouragement groups had purchased any E-tag/Ag-verify products and had purchased E-tag herbicide specifically in first season 2017, but very few had purchased E-tag or any herbicide in second season 2016. It appears that very few people purchased E-tag maize at all, especially in second season 2016. This is likely because farmers tend to purchase their inputs in the first growing season of the year and then continue to use them (if they are still available during the second growing season).

Turning now to the encouragement interventions in first season 2017 (in second season 2016 the patterns are mirrored but the numbers are slightly lower), almost no households attended the community promotion meeting in the control group, and about 40 percent of households attended the meeting in the encouragement and discount groups, with slightly more households from the discount group attending. Approximately 40 percent of those who attended the meetings found the meetings helpful. Almost no households in the control group reported receiving in interactive voice messages as well, and about 28 percent of the encouragement group and 31 percent of the discount group reported receiving the voice messages. Further, 19 percent of the encouragement group and 20 percent of the discount group interacted with the voice messages.

Looking at the discounts that were only distributed in first season 2017, almost nobody in the control group received the discount vouchers, 16 percent of people in the encouragement group (includes villages not assigned to the discount) received the vouchers, and about 30 percent of people received the maize and herbicide vouchers (this is not surprising as they were both given out together). Based on data collected on the voucher program during the CPMs, there was a modest use of the vouchers by those that received them. On average, 55 people per community assigned to the discount treatment received a voucher for herbicide and a voucher for hybrid maize. In these communities, on average 17 people redeemed the herbicide voucher for a rebate and 10 people redeemed the hybrid maize voucher for a rebate. This is consistent with the overall usage patterns; households are much more likely to use herbicide compared to hybrid maize.

Unfortunately, as has been found with other similar authentication systems, very few farmers authenticated their E-tag herbicide or hybrid maize. Less than one percent of the control and encouragement groups attempted to authenticate E-tag maize, and only 2 percent of the discount group attempted. For E-tag herbicide, almost nobody in the control group attempted to authenticate in the control group, while 10 percent of people in the encouragement group and 15 percent of people in the discount group attempted to authenticate. In the discount group, enumerators often helped with authentication during the time when the discount vouchers were redeemed. We see similarly low rates of success in authenticating E-tag maize across groups, but we do see that 14 percent of the discount group was successfully able to authenticate their E-tag herbicide.

Data from the IVR messaging system shows that of the phone numbers that we were able to list from the CLE samples, more than 90 percent of attempted calls went through. People also listened to the bulk of messages, with approximately one third of respondents completing the message to the end, and an average call duration of two minutes. Over 40,500 calls were made in total over two growing seasons informing farmers about tagged products and letting them know when the products were available in their local markets.

Impact Estimates

Here, we discuss the report's main findings on the causal impact of the e-verification program. We find that the e-verification program led to a broad increase in take-up of agricultural inputs in first season 2017, the season in which the e-verification program was most operating most intensively. The program caused a statistically significant increase in the use of hybrid maize seed by 5.4 percentage points. This effect is large given that only 10.5 percent of households used hybrid maize seed in the control group at endline. The impact on use of any maize seed is an increase by 5.6 percentage points. E-verification caused an increase in take-up of glyphosate herbicide of 7.7 percentage points (compared to a 43.8 percent adoption rate in the control group) and of any herbicide by 6.6 percentage points. Use of inorganic fertilizer increased 3.2 percentage points, but this effect was only weakly significant. Use of any fertilizer increased 5.3 percent. These effects on fertilizer represent a spillover effect of the program because fertilizer was not included in the Kakasa e-verification program. These results are very positive; the e-verification program caused a broad increase in the adoption of high-quality agricultural inputs, thereby showing that such a verification scheme has potential to foster technology adoption.

In second season 2016, when the e-verification platform was still very new, effects on take-up of high-quality inputs were limited, with an increase in use of hybrid maize by 3.3 percentage points and of any maize of 3.7 percentage points (weakly significant). There were no significant impacts on take-up of other inputs in that season. These weaker results likely reflect that second season is a much less important growing season in Uganda and the e-verification scheme was very new.

There were far fewer impacts on other farm outputs include maize yields and net income from farming. In first season 2017, there were no positive impacts of the program on number of crops grown, maize yield, log value of farm production or net income from farming. A modest reduction in net income from farming was weakly significant. In second season 2016, e-verification increases the number of crops grown by 0.18 crops; there were no impacts on any other input in that season. However, we are not surprised that there were limited statistically significant results on productivity and farm income. Prices, land area, and volumes of inputs purchased and harvest difficult for farmers to remember accurately, so there is a lot of measurement error in these variables. Once these variables are aggregated to form a measure of farm income, for example, the measure is quite noisy. In addition, the economic benefits of these modern inputs may accumulate over several seasons as farmers learn more about the inputs. In addition, the main benefit of glyphosate herbicide is in reducing the amount of labor time spent weeding (Ashour et al., 2017), which may affect farm income but would be less likely to affect yield.

We examined the impact of e-verification on farmers' beliefs or perceptions about the extent of counterfeiting and adulteration of agricultural inputs in the market. Having an effect on farmers' beliefs about the extent of counterfeiting and adulteration is an important step in building confidence with consumers in the quality of inputs in their markets. E-verification led farmers to believe that it was more likely that maize (conventional or hybrid) quality was ever lowered by counterfeiting/adulteration, by 10 and 16 percentage points, respectively, and these coefficients are statistically significant. These differences in beliefs represent an 18 and 22 percent change, respectively, which are quite substantial effect sizes. We find no effect of e-verification on beliefs about whether inorganic fertilizer or glyphosate herbicide are ever counterfeited or adulterated; point estimates on inorganic fertilizer and glyphosate herbicide are negative but are not statistically significant. Thus, e-verification shifted farmers' beliefs about the quality of the agricultural inputs in their markets – this is an important result since beliefs are integral to adoption decisions.

We are able to further understand patterns of farmer beliefs by separating beliefs regarding Kakasa and non-Kakasa inputs. We asked farmers, for each input (hybrid maize, glyphosate herbicide, and inorganic fertilizer), whether none, less than half, half, more than half, or all is counterfeit/adulterated, and we did so separately for Kakasa and non-Kakasa products. For Kakasa hybrid maize and herbicide, farmers report that it is less likely that half or more of the input is counterfeit/adulterated by 13 and 5 percentage points, respectively. Also, e-verification led to an increase in beliefs that more than half of non-Kakasa glyphosate herbicide is ever counterfeited or adulterated. These results show that farmers had confidence in the quality of the tagged products and suggest that the shift in beliefs was due to relatively greater confidence in tagged products compared to products that were not tagged.

Next, we explored whether there was heterogeneity in the impacts of e-verification: that is, were there differences in impacts of the program along important dimensions of farmer characteristics. We separately identify impacts for male and female primary agricultural decision-makers (PADMs); PADMs who have completed primary school or more and PADMs who have not completed primary school; households considered to be poor (belonging to the lowest quintile of an asset index) and not poor; PADMs who are more or less risk-averse; PADMs who are ambiguity averse and not ambiguity averse, and for PADMs who have a phone or do not have a phone. For each of these characteristics, we examine heterogeneity of impacts on take up of inputs as well as farming outcomes. We find that impacts of e-verification on take-up of hybrid maize and inorganic fertilizer both appear to be driven by increased adoption by men. Regarding education, there does not appear to be a systematic difference in the impact of e-verification based on the education level of the PADM. Regarding poverty, the impacts of the intervention are statistically significant for take-up of all inputs for the non-poor group (compared to the non-poor control group), but there are no statistically significant differences in impacts between the poor and non-poor groups. In the poorest quintile, only impacts on hybrid maize are statistically significant; weaker effects for the other inputs sometimes reflects low power due to the small sample. Also, E-verification caused households in the poorest quintile of the asset index to increase take-up of hybrid maize seed by 15.3 percentage points from only 2.4 percentage points in the control group. Impacts tended to be larger for individuals owning or having access to a phone. And there are no systematic differences in impacts by preferences regarding risk aversion and ambiguity aversion.

Finally, we introduced a system of input vouchers in first season 2017 targeted to a randomly selected encouragement communities to better learn about how an experimentally targeted discount could increase take-up and thereby reduce counterfeiting in the sense of raising average quality. The random assignment is thus at the market location level, since it is market locations that stock the inputs where we would be able to measure differences in quality. Input quality was measured as the genetic distance from seed samples collected in the field to reference seed samples for hybrid seed and as the measured concentration to the claimed concentration (on the package label) of glyphosate for herbicide. For hybrid maize seed, we do not find any significant impact of e-verification on average quality of hybrid maize seed in the market. However, estimates are suggestive of better quality of hybrid maize in discount communities and that this association was larger for non-Kasasa maize. This result documents an association in the data that is consistent with the objectives of the e-verification program – to reduce the prevalence of low-quality inputs outside the E-tag scheme by increasing demand for E-tag products and making lower quality products outside the scheme less profitable. For glyphosate herbicide, the measure of input quality is very noisy, in part due to a long period of storage before samples could be tested because we did not receive approval to export the samples for more than 1.5 years. Thus, we cannot conclude that e-verification improved quality of herbicide.

Conclusion and Recommendations

The encouragement intervention lead to large differences in knowledge of the E-tag and Ag-verify verification schemes, of over 50 percentage points in first season 2017 and second season 2016. The experimental encouragement program was effective at creating a wedge between high take-up encouragement treatment villages and low take-up encouragement control villages. The model then uses differences in the predicted probability in knowledge about E-tag and Ag-verify between encouragement treatment and control villages to estimate the effect of e-verification on agricultural input use and farm productivity and profit. The results show that the e-verification initiatives were remarkably effective. In first season 2017, when e-verification was largely scaled up in the study area, the program caused large increases in take-up of agricultural inputs, including a 5.4 percent increase in use of hybrid maize seed; a 7.7 percent increase in use of glyphosate herbicide; and a 5.3 percent increase in use of both inorganic fertilizer and any fertilizer. These are large effects, substantially increasing take up of hybrid maize seed and inorganic fertilizer, especially.

In first season 2017, there are no statistically significant impacts on farm outcomes such as yields or net income. This is mainly due to the noisy nature of the data collected. In second season 2016, however, we are able to detect a significant increase in the number of crops grown in the encouragement group.

The e-verification scheme led to an increase in beliefs about whether the quality of conventional maize seed or hybrid maize seed is sometimes lowered in the market, which may reflect the increase in awareness about the presence of the e-verification authentication scheme. Also, impacts of e-verification on the input use appears to be driven mostly by men in the sample. Impacts on use of hybrid maize seed were somewhat large for the poorest (based on asset quintile) households in the sample, but this adoption effect for the poor did not lead to positive impacts on yield or net income for poor households. Finally, there were no statistically significant impacts of e-verification on the quality of maize seed or herbicide in the market (a measure of counterfeiting), though associations with maize seed quality suggests that it is plausible that the scheme could eventually lead to changes in the input market that would drive out low quality producers.

Based on these findings, we have the following recommendations. First, the e-verification scheme should be expanded to promote further increases in adoption of high-quality inputs. In addition, E-tag and Ag-verify are currently designed to provide *product assurance* (what you see on the label is what you get), but not *quality assurance*, as there is no testing or enforcement mechanism to assure that the products are actually of high quality. E-verification should be expanded to include such testing and enforcement components to further increase confidence in input markets and foster greater profitability for both input companies and farmers.

CONTENTS

List of Tables

List of Figures

Acknowledgments

We gratefully acknowledge funding for this evaluation study from USAID through agreement number AID-BFS-IO-14-00002. We have received extremely helpful advice and guidance about agricultural inputs in Uganda and the counterfeiting problem from Bruce Kisitu at the Grameen Foundation. We are also grateful for help and support from Andrew McCown, Corey Fortin, Martin Fowler, May Mwaka, Peter Birigenda, and Andrew McKim at the USAID Kampala mission and from Tatiana Pulido and Emily Hogue at the USAID Bureau for Food Security. We also received helpful advice from TetraTech staff, including from Eric Derks and Rita Laker-Ojok, Chiefs of Party of the FTF Agricultural Inputs Activity, and from Andrew Gita. We had many helpful conversations with Rudolf Guyer, Director General of Crop Life Africa and Middle East, about the design and potential rollout of an e-verification system for agricultural inputs in Uganda. We further thank the Ugandan Government Analytic Laboratory for conducting testing of one round of herbicide samples, and Don Cooper for conducting testing on the second round of herbicide samples and providing invaluable advice. We also thank the LGC Genomics laboratory for conducting testing on hybrid maize seed samples, as well as Dr. Julius Sserumago for the analysis of these data. We also thank Jessica Hoel for advice through many helpful conversations about this project. Thanks to our talented survey field team lead by Geoffrey Kiguli, Fieldwork Coordinator, and Fiona Namugenyi and Jaspher Okello, Finance and Logistics Coordinators. We are also grateful to all of the survey respondents for taking the time to respond to the survey and tell us their stories.

This publication was made possible through support provided by the Bureau for Food Security, U.S. Agency for International Development, under the terms of Award No. AID-BFS-IO-14-00002. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Agency for International Development.

This work was undertaken as part of the CGIAR Research Programs on Policies, Institutions, and Markets (PIM) and Agriculture for Nutrition and Health (A4NH). Funding support for this study was provided by USAID. This publication has not gone through IFPRI's standard peer-review procedure. The opinions expressed here belong to the authors, and do not necessarily reflect those of PIM, A4NH, IFPRI, or CGIAR.

Acronyms

BMGF	Bill and Melinda Gates Foundation
CAPI	Computer assisted personal interview
CLE	Community listing exercise
CPM	Community Promotion Meeting
EV	E-verification
FTF	Feed the Future
GPS	Global Positioning System
ID	Identification
IFPRI	International Food Policy Research Institute
IR	Intermediate Results
IVR	Interactive Voice Recording
LC1	Local Council 1 (community)
MDE	Minimum detectable effect
MH	Market hub
ML	Market location
PADM	Primary agricultural decisionmaker
RCT	Randomized controlled trial
RMP	Representative maize plot
SADM	Secondary agricultural decisionmaker
SMS	Short Message Service
UNBS	Uganda National Bureau of Standards
USAID	United States Agency for International Development

1. Introduction

1.1 Motivation

Across Africa South of the Sahara, the take up of productivity- and income-enhancing inputs such as fertilizer, herbicides, and enhanced varieties of seeds is extremely low (Aker, 2011; Kelly et al., 2003; Duflo et al., 2011; Morris et al., 2007; Bold et al., 2015). This pattern is stark despite the importance of farming in the region. More than seventy percent Africans living under US$1.25 per day are engaged in small-holder farming (IFAD, 2011; Foster and Rosenzweig, 2010). Improving productivity on these farms is thus essential to reducing poverty rates and improving food security, as well as numerous other outcomes (Hirvonen and Hoddinott, 2016; Byerlee et al., 2009; Ligon and Sadoulet, 2008; Bravo-Ortega and Lederman, 2005; Ravallion and Chen, 2007; Irz et al., 2001; FAO, 2009).

In Uganda, use of high-quality agricultural inputs like hybrid maize seed, herbicides, pesticides, and fertilizer is extremely low (Ashour et al, 2015). This limits incomes and leads to low agricultural productivity that continues to be hampered by poor agronomic practices; low quality germplasm; declining soil fertility; and losses due to pests, disease, and postharvest handling practices. The negative effect of these factors on agricultural technology adoption have been compounded by a lack of farmer trust in the current inputs supply system, which has been plagued by counterfeiting (Fairbairn et al., 2018; Tjenstrom et al., 2017; Kilic et al., 2017; Bold et al., 2018).[2]

Although there is growing evidence on rates of counterfeiting, the perception is that counterfeiting is very common, and this depresses demand for these inputs, reduces prices, and increases risk for farmers (Ashour et al, 2018). Counterfeit products range from benign fake or adulterated materials, to banned substances that are harmful to both crops and human health. Counterfeiters have become increasingly innovative in their techniques, making it difficult to identify their products without laboratory tests, while farmers and agro-dealers have little means of verifying whether a product is genuine, unexpired, priced fairly, or accurately labeled by brand, type, or concentration. Without significant investments in high-quality inputs, current domestic production levels will not be sufficient to support Uganda's growing population nor its plans to increase exports to support regional food security.

[2] Tampering with agricultural inputs can take the form of adulteration, where a product such as herbicide is mixed with an inexpensive inert ingredient like water, or complete counterfeiting in which the product is replaced in its packaging with a fake. Throughout this report, we will refer to each of these as counterfeiting.

1.2 E-verification: Addressing counterfeiting through product assurance labels

To help mitigate this issue, USAID through the Feed the Future Agricultural Inputs (FTF Ag Inputs) activity has supported a program for input product assurance called e-verification (EV). E-verification involves labeling agricultural inputs with a scratch-off label that provides an authentication code that can be used to confirm that the labeled product is what is claimed on the packaging (for example, brand, bottle size). The consumer sends the code that they have revealed under the scratched off label to a short code and receives back an SMS message confirming the identity of the product. A pilot version of this approach was undertaken in 2012 for herbicides with the support of Crop Life Uganda and Crop Life Africa and Middle East. The pilot demonstrated that there was significant demand for e-verified herbicide and that farmers were willing to pay a modest price premium for this form of product assurance. The new USAID project supported a scaled-up e-verification initiative led by Tetra Tech under its FTF Ag Inputs activity. Given the potential importance of this initiative, USAID selected the International Food Policy Research Institute (IFPRI) to conduct an independent impact evaluation of the effectiveness of the EV system in improving adoption of high-quality inputs and reducing the prevalence of counterfeiting.

1.3 Objectives

The objectives of the e-verification project are to increase adoption of productivity-enhancing agricultural inputs, to increase farmer profits and household welfare as a result of adoption, and ultimately, to reduce the prevalence of counterfeit and adulterated agricultural inputs on the market (this particular outcome will be the subject of a separate report). Using a rigorous experimental design, the IFPRI impact evaluation estimates the impact of the e-verification scheme on adoption of high quality agricultural inputs, agricultural productivity, farm profits and the prevalence of counterfeiting. The evaluation will also show how prices for agricultural inputs contribute to low adoption in the face of the uncertainty caused by input counterfeiting. The three main agricultural inputs that we study are hybrid maize seed, glyphosate herbicide, and inorganic fertilizer.

This impact evaluation addresses the following primary research questions:

1. Does improving access to verified agricultural inputs result in greater take-up of high-quality inputs by farmers?
2. What is the impact of increased use of agricultural inputs on the gross margins and yield for Ugandan farmers?
3. Do impact estimates differ by farmer characteristics such as gender, income, and others?
4. Does improving access to verified agricultural inputs result in a lower prevalence of low quality agricultural inputs?

The baseline report for this study introduced the randomized encouragement study design, summarized characteristics of sample households and farms, and presented evidence that these characteristics were balanced between the encouragement treatment and control groups (Ashour et al. 2015).

This endline report summarizes the impact evaluation study design, describes the sample, and provides impact estimates for the effect of e-verification on the outcomes listed above. Section 2 of this report introduces the FTF e-verification sub-activity and provides details on how e-verification works. Section 3 summarizes the randomized encouragement impact evaluation design and the details of the interventions. Section 4 describes the sampling strategy, Section 5 provides details of the data collected through the project. Section 6 explains the strategies we use to identify causal impacts of the program. Section 7 provides the impact estimates for the first two research questions. The impact estimates for the third research question are presented in Section 8, and Section 9 concludes with some discussion.

2. The E-verification Sub-activity

2.1 Background

The objective of the EV initiative in the FTF Ag Inputs activity is to increase production of maize, beans, and coffee through the appropriate use of high-quality agro-inputs (seeds, fertilizers, herbicides, and pesticides). The FTF Ag Inputs activity, which began in 2013, was implemented in 15 selected FTF target districts. The activity operated until mid-2018. The agricultural inputs considered under this impact evaluation are hybrid maize seed, glyphosate herbicide, and inorganic fertilizer.

The impact evaluation will focus on ascertaining the impact of two of the four Intermediate Results (IR 1 and 2) of USAID/Uganda's Agriculture Inputs Activity:

- IR1 "Increased availability of high-quality inputs to farmers in FTF focus districts"
- IR2 "Decreased prevalence of counterfeit agricultural inputs"

In terms of FTF objectives, the project relates to those of improving agricultural productivity and increasing private investment in agriculture. The impact evaluation tests the following hypothesis: "If USAID/Uganda introduces and promotes electronic verification (EV) and effective marketing of agricultural inputs, then Ugandan farmers in key FTF districts will demonstrate higher adoption rates of those inputs, resulting in a commensurate improvement in agricultural productivity."

Building on the experiences of other promising systems piloted in Uganda and elsewhere, the FTF Ag Inputs activity undertook the development of technical, regulatory, and management infrastructure needed to introduce a new e-verification system of product assurance for agricultural inputs. The EV system added labels to packages of inputs. These labels will link the input package to the SMS-based product assurance system. Codes printed on the labels will provide users with information such as brand, input type and concentration, batch number, date of manufacture, and expiry. EV packaging will instruct consumers purchasing the inputs to "authenticate" the product by sending an SMS message to the EV system, entering the package-specific code, and receiving an SMS message in return verifying that the product is genuine.

2.2 E-tag, Ag-Verify, and CHK

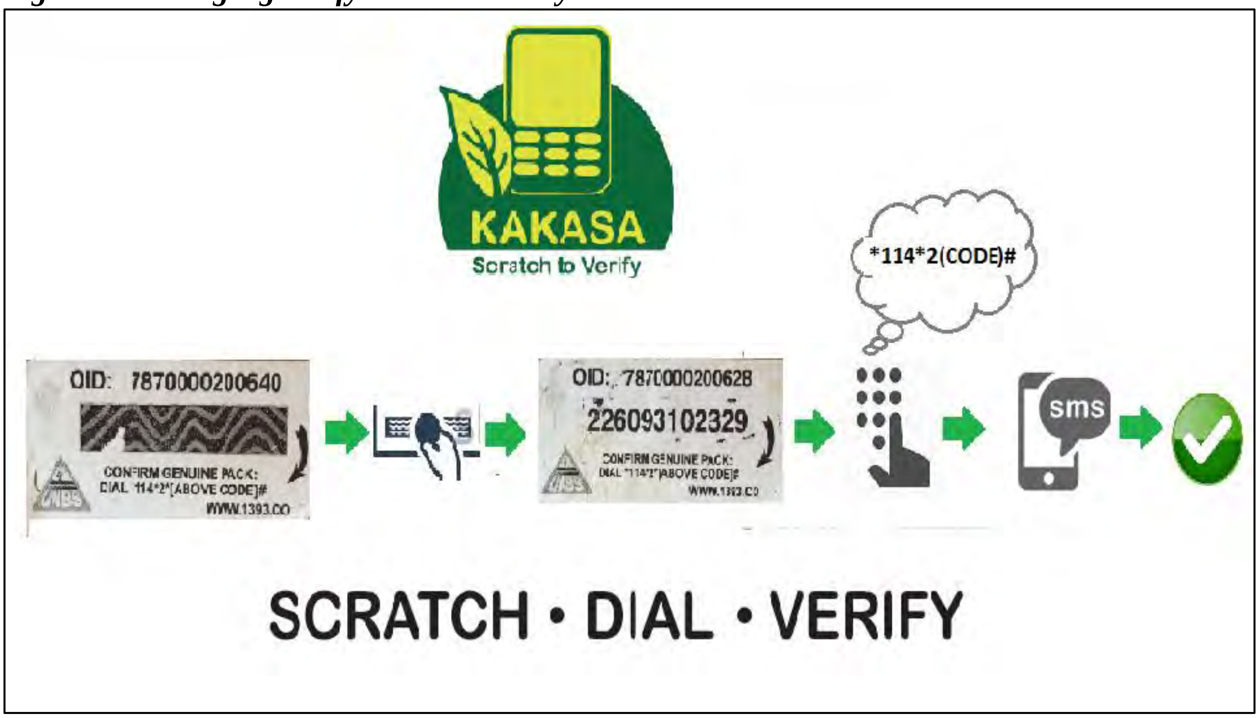
Initially, AgInputs was in discussion with numerous actors including the Bill and Melinda Gates Foundation (BMGF), Crop Life Africa and the Middle East and Crop Live Uganda, among others. However, the Uganda National Bureau of Standards (UNBS), which has the mandate to ensure quality standards are met nationally across a wide array of products, ultimately implemented the scheme under the name 'E-tag', also known as 'Kakasa'. The UNBS labels were cheaper and were embedded in the government, which fully supported this activity. Thus, many companies bought into the 'E-tag' scheme. The scheme was directed at glyphosate herbicide and hybrid maize seeds (with a small market share of beans and other crops). Fertilizer has not yet entered the scheme. During 2017, E-tag was supplemented by another scratch code scheme that also fell under UNBS called 'Ag-verify', which was specifically for hybrid maize seed.

The E-tag/Ag-verify systems worked as follows:

1. Stick-on labels were applied to bottles of herbicide or bags of maize by the company
2. When farmers purchased an E-tag/Ag-verify product, they were to authenticate the product. They would do so by scratching of the scratch panel on the label and texting the scratched code to a short code.
3. The farmer would receive a message back with details of the product.

Figure 2.1 displays the process graphically.

Figure 2.1. E-tag/Ag-verify scratch label system



The E-tag scheme was taken up by Bukoola (for their 1 litre and half litre bottles of Weed Master herbicide) and by Balton (for their 1 litre bottle of Roundup) and was run in conjunction with the SMS platform of mPedigree. The Ag-Verify system was taken up by Otis Seeds, Pearl Seeds, Naseco, Grow More Seeds, Simlaw Seeds, Equator Seeds, FICA, and Masindi Seeds for their hybrid varieties. Each tag costed 60 UGX and over 100,000 labels were distributed in 2016 alone.

The Kakasa advertising campaign was extensive. It included road shows, brochures and ads in store selling agricultural inputs, radio advertisements and segments, and even television segments.

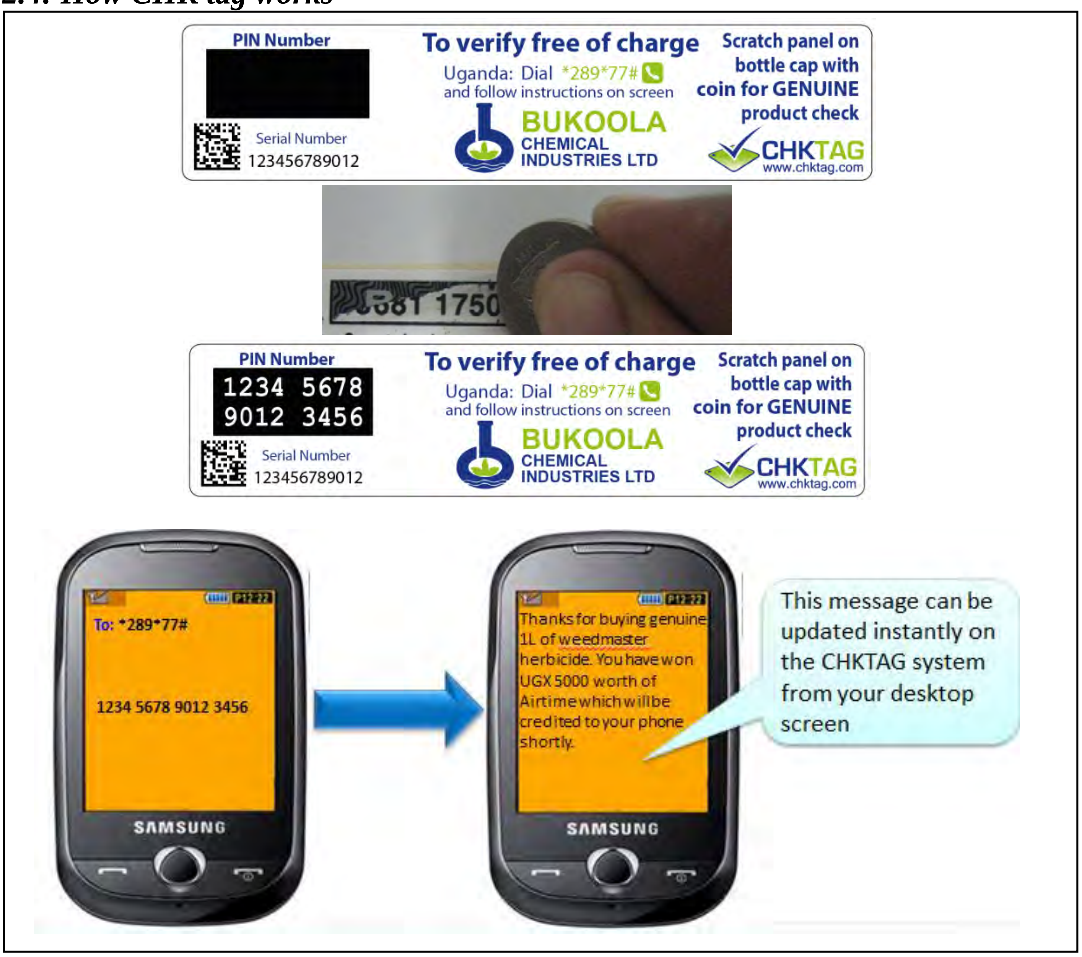
It was talked about on popular shows shown in public transportation and reached farmers quite broadly. There was also a call centre for people to report fake inputs. Figure 2.2 below shows an advertising poster for the Kakasa campaign.

Figure 2.2: Kakasa Poster

Another scheme called CHK tag was provided by Brand-ID and was also taken up by Bukoola for their 1 litre bottles of Weed Master. CHK's tag looked slightly different, but the process was analogous to that for E-tag. Figure 2.3 shows what the CHK label looked like and Figure 2.4 shows how the CHK authentication process worked.

Figure 2.3: CHK tag label

Figure 2.4: How CHK tag works



3. The Impact Evaluation Design

3.1 Theory of change

This project tests the following hypothesis: counterfeiting and adulteration of agricultural inputs is a major impediment to promoting adoption of high-quality inputs that impedes growth in agricultural productivity and farm incomes. If counterfeiting and adulteration are significant impediments to the adoption of high-quality inputs, then providing a product that is guaranteed to be genuine should improve the take-up of high-quality inputs.[3]

If farmers suspect that inputs may be fake or adulterated, this causes them to reduce their demand for the products. Farmers who would otherwise use the products may avoid them altogether. Reduced demand arises because the uncertainty about the products' authenticity reduces farmers' expected returns from adopting the products. This lack of demand reduces input prices and harms the profitability of selling genuine high-quality inputs. As a result, many such genuine inputs may cease to be available in the market, leaving only inputs of suspect quality or low quality. This problem of missing information about a product harming demand is known in economics as "adverse selection." The counterfeiting problem is a form of the adverse selection problem described by Akerlof (1970) in his paper on the effect of unobserved quality of used cars, "The Market for `Lemons'." If the quality of remaining products on the market becomes known to farmers through experience, prices may adjust to reflect the lower returns to using these counterfeit products (some versions provide some productivity boost), but this still reduces availability of higher quality, productive technology that could raise incomes.

The data in this study show that farmers do believe that many agricultural inputs are fake (Ashour et al, 2018). If farmer beliefs about the quality of inputs align well with the actual rates of counterfeiting in markets, this would suggest that such a verification scheme would be welfare improving. However, if farmer beliefs do not align well with actual rates of counterfeiting, this would suggest that an information campaign would be most useful. Ashour et al (2018) find that farmer beliefs do align with actual rates of counterfeiting, but that these beliefs are imperfect. Thus, a verification system and an information campaign could both be useful.

The EV sub-activity could help to overcome this problem by providing a form of product assurance with information for agricultural inputs. This would help to increase demand for those inputs. Improved take-up of genuine productivity-enhancing inputs should help in increasing yields, farm productivity, and household income. This impact pathway assumes that the improved inputs increase yields and farm productivity (i.e. that farmers make the necessary complementary investments required for these inputs to be productive), and that improved productivity increases household welfare (so farmers do not then substitute away from other activities, leaving total

[3] Further details on the theory of change are provided in the Baseline Report for this study (Ashour et al, 2015).

welfare unchanged). Consumers of farm products may also benefit through lower prices for crops that use high-quality inputs.

3.2 Evaluation Design

The impact evaluation of the EV component to the FTF Ag Inputs activity uses an encouragement design to causally identify the impacts of access to e-verified inputs on take-up of high-quality inputs, prevalence of counterfeiting, farmers' yield and gross margins of maize, and household welfare.

Markets gain access to agricultural products through a variety of conduits, including directly from suppliers, through distributors reaching out to retail markets, and by retailers visiting distributors based in major cities. This makes it impossible to systemically assign retail markets into treatment and control groups, thus ruling out randomization as a method for identifying impacts.

Several studies in the economics literature have used encouragement designs to identify the impacts of interventions that would not otherwise be possible using RCTs. Studying adoption (be it agricultural inputs or other programs) is particularly well-suited to this approach. When products or programs are available to everyone, random variation can be induced by encouraging some people to adopt and not others. Duflo and Saez (2003) study enrollment in a retirement savings plan in a large university. They use reminder letters as well as financial incentives to randomly selected individuals in randomly-chosen departments to attend an information seminar on retirement savings. They find that both people who received the letter and incentive, as well as others in the same department as the person receiving the letter, were more likely to enroll in the program. Duflo, Kremer, and Robinson (2011) use free delivery of fertilizer (to reduce transactions costs) to encourage farmers in Kenya to adopt this input. They find significantly higher adoption in the group of farmers receiving the encouragement, despite the fact that not everyone offered the free delivery makes use of it. Thornton (2008) finds that in rural Malawi, monetary incentives to learn one's HIV status doubles the share of those who seek this information. Finally, Katz, Kling, and Liebman (2001) use randomly allocated housing vouchers to study the effects of neighborhood changes. They find positive impacts on safety, health, and behavioral problems among boys.

By randomly assigning one of two matched communities in each market location to receive 'encouragements' – in our case interactive voice recording (IVR) messages and community promotion meetings (CPMs) – the encouragement design produces exogenous variation in the rate of take-up of e-verified products, with higher take-up, on average, in communities receiving the encouragement. Differences in average rates of take-up of e-verified products between treatment and control villages provide a measure of the effectiveness of the encouragement treatment. However, the encouragement design can also be used to estimate the impact of access to e-verification on take-up of high-quality inputs, taking advantage of the experimentally-induced variation in take-up of e-verification. As we explain more fully below in Section 6, in a regression model of the impact of e-verification on adoption of high-quality inputs like hybrid maize seed, the encouragement assignment variable will be used as an instrumental variable for the purchase of e-verified products, providing an unbiased estimate of the causal impact of e-verification on adoption of agricultural inputs. Two rounds of IVR messaging and CPMs were conducted: the first was in the second growing season of 2016 and the second was in the first growing season of 2017.

We also varied the price of the e-verified product through CMP promotion campaigns. During the second CPM meetings, attendees were provided with vouchers to receive a discount on E-tag herbicide and E-tag/Ag-verify hybrid maize. Importantly, the discounts were randomly assigned at the market location level. As stated above, in 120 market locations, 120 villages (of 240) were assigned to the control group, and 120 villages received encouragement. Of the 120 encouragement villages, 60 villages received the two vouchers to purchase E-tag or Ag-verify herbicide or hybrid maize. The impact of the encouragement on the take-up of EV products, and the impact of EV on take-up of agricultural inputs will be identified from the differences between the 120 control and 120 treatment communities, see Table 3.1 below.

Table 3.1: Treatment arm design across 240 villages

	Encouragement	Control
No discount	60	120
Discount	60	

3.2.1 The encouragement IVR messages

We used a web-based platform called 'Voto Mobile' to send the IVR messages and track whether messages were sent and received, and can be used to export these data into Microsoft Excel files. The research assistant in Washington, coordinating with the research team, managed encouragement IVR messages. This platform allowed the grouping of phone numbers into both locations and 'statuses'.

Two different types of IVR messages were sent to farmers in each round of messaging depending on what the status of the market location was. First, farmers would receive a message that Kakasa products would soon be available in their local markets. The messages explained what Kakasa products are, why they exist, how to use them, and a reminder always to authenticate. The message then gave the listener the option to press 1 to hear a list of Kakasa herbicide that would be available, and to press 2 to hear a list of Kakasa maize that would be available. Options were also available to repeat messages, and to call the number back (which was free). This message was attempted three times; if there was no answer or if the phone was off, there were then still two more chances for the individual to receive the phone call.

The second IVR message was sent once products were available at the farmers' local markets. To determine availability of Kakasa inputs in particular markets, a system of 'market monitoring' was set up. A team of enumerators in the IFPRI office in Kampala phoned shop owners in every market location once a week to see if they had stocked any Kakasa products. Once a shop stated that they did stock Kakasa products, the status of the market location was changed to having stocked Kakasa products, and the second IVR message was sent to farmers informing farmers that Kakasa products were available right now in their local markets. Once again, they had the option to press 1 for the list of herbicide products, and 2 for the list of maize products, and they could also repeat the messages and phone the number back. This message was also attempted three times.

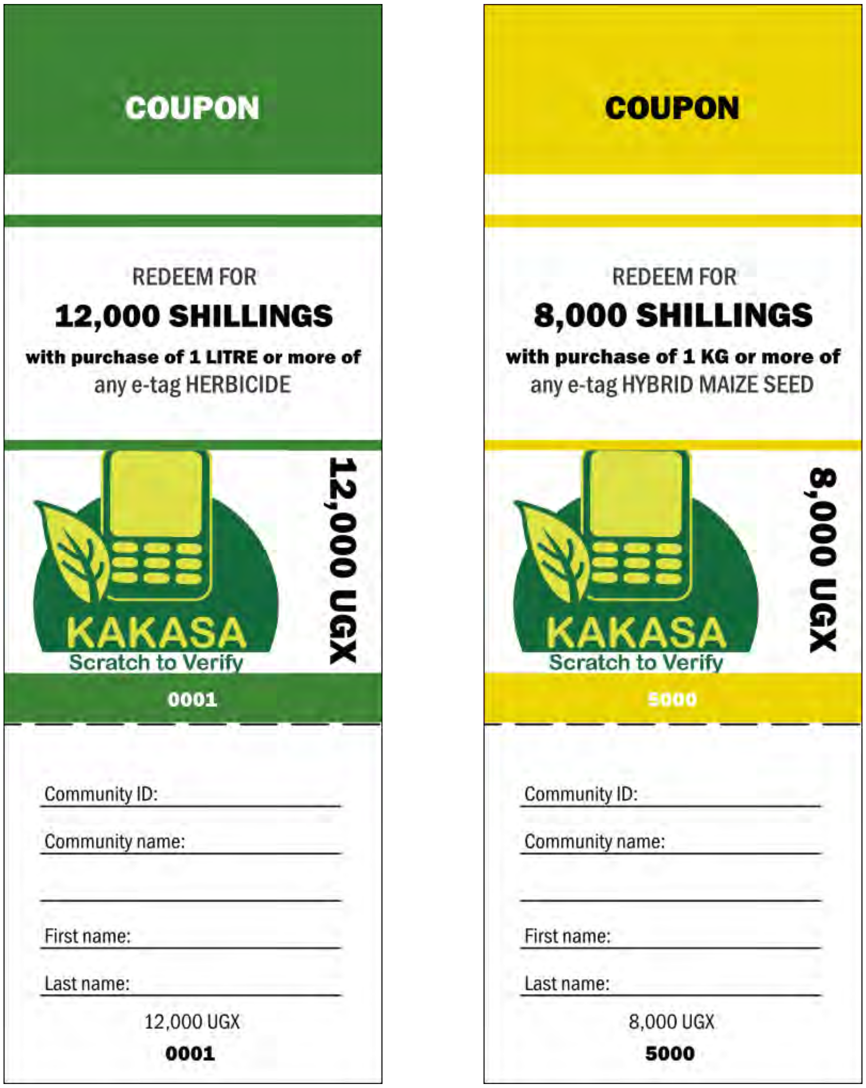
3.2.2 Community Promotion Meetings

The community promotion meetings also took place during the second growing season of 2016 and the first growing season of 2017. They generally occurred in between the first and second IVR messages. A team of two enumerators visited all 120 encouragement villages. Two days before the scheduled meeting, the enumerators would phone the village leader to inform them that the meeting would occur and would enlist their help in spreading the word **only** in that particular village. When the enumerators arrived in the village, they would first go around the village with a guide and inform people that a meeting would be held shortly that would introduce a new product and that prizes could be won so as to encourage people to join the meetings. Meetings were held in a centrally located place within the community. Upbeat music was played while people arrived, and raffle tickets were handed out on arrival. The village leader provided an introduction to the team, and a skit was performed by the pair of enumerators. The enumerators then conducted a demonstration of Kakasa products. There was a question-answer session where correct answers would win people prizes, and there was a raffle to give away Kakasa products.

3.2.3 Discount Vouchers

During the second round of CPMs in the first season of 2017, vouchers were handed out at the end of the meeting. Meeting participants were told that each person attending the meeting would receive two vouchers: one for purchasing Kakasa herbicide and one for purchasing Kakasa maize. Figure 3.2 shows the vouchers that were handed out. The herbicide coupon was worth 12,000 UGX, which most often covered the entire price of a 1 litre bottle of herbicide (the average price being 12,000 UGX), and the maize coupon covered the bulk of the cost of 1 kg of herbicide (the average price being just over 8,000 UGX). The vouchers were specially printed to avoid counterfeiting of the vouchers. They were printed on a special type of paper whereby rubbing the back of the paper causes the faint words written there to disappear for a few moments and then re-appear. This type of paper was only available in the US and not in Uganda, so we are confident that the vouchers were not counterfeited. Vouchers could be split into two parts (on the dashed line). Community members kept the top part, and the enumerators kept the bottom part. On the bottom, the names of the voucher recipients and the community IDs were kept so as to keep track of the vouchers given out. Both parts had a unique code identifying the voucher number. Community members were informed that the two enumerators would return to the community in approximately three weeks to provide redemptions for the vouchers. In order to redeem a voucher, the individual needed to provide the original voucher, and proof of purchase (this could be a receipt, an empty container, or a full container). If the tag was not scratched to authenticate, the enumerator would do so with them upon redemption of the voucher. Enumerators kept track of how many vouchers were given out in each community and how many were redeemed. Before they returned to the village, the village leader was informed of the day that they would be arriving.

Figure 3.1: Discount Vouchers

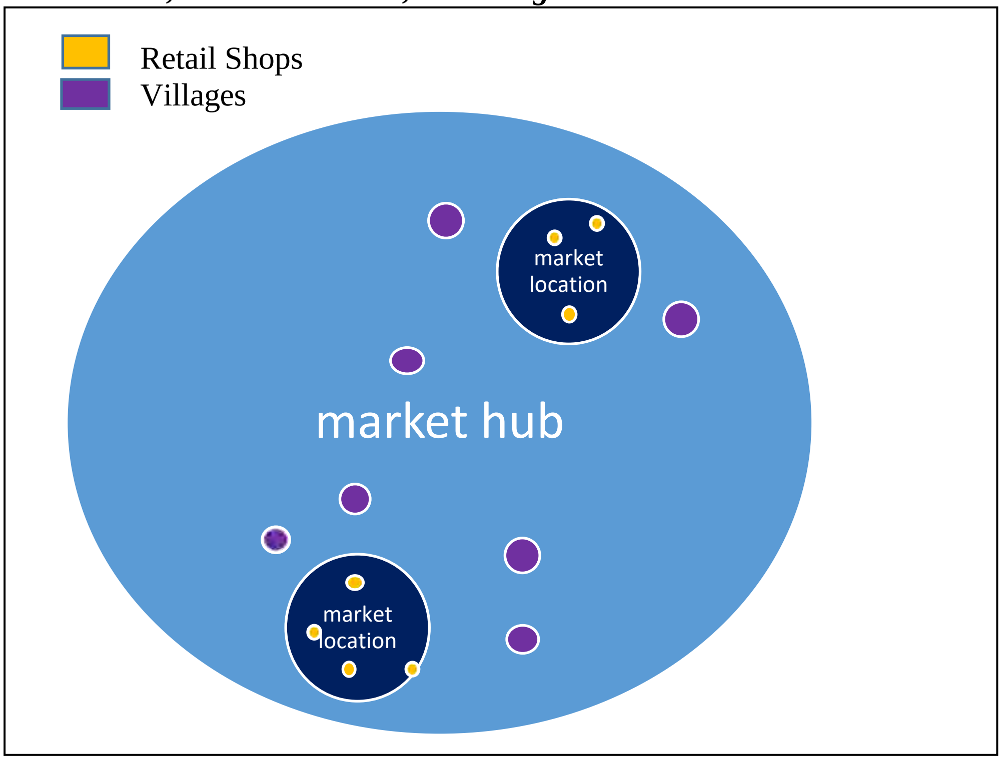


4. Sample Design

4.1 Site selection and sampling

The first step of the sampling process was identifying broad areas in which to work. IFPRI worked with Tetra Tech to identify “market hubs” where e-verified products were likely to be made available. A market hub is a major market area consisting of a collection of “market locations” covering one or more districts, centered around a major town (generally, the district town center). A market location is a collection of retail shops (often referred to as a trading center) selling agricultural inputs in the same vicinity. Each market location serves several surrounding villages (see Figure 4.1). The most common method for households from these villages to access agricultural inputs is from retail shops in their nearby market location, although they may also travel to the market hub town to source their inputs. Less common would be sourcing inputs directly from a distributor or traveling greater distances, to other market hubs or to Kampala, to purchase agricultural inputs.

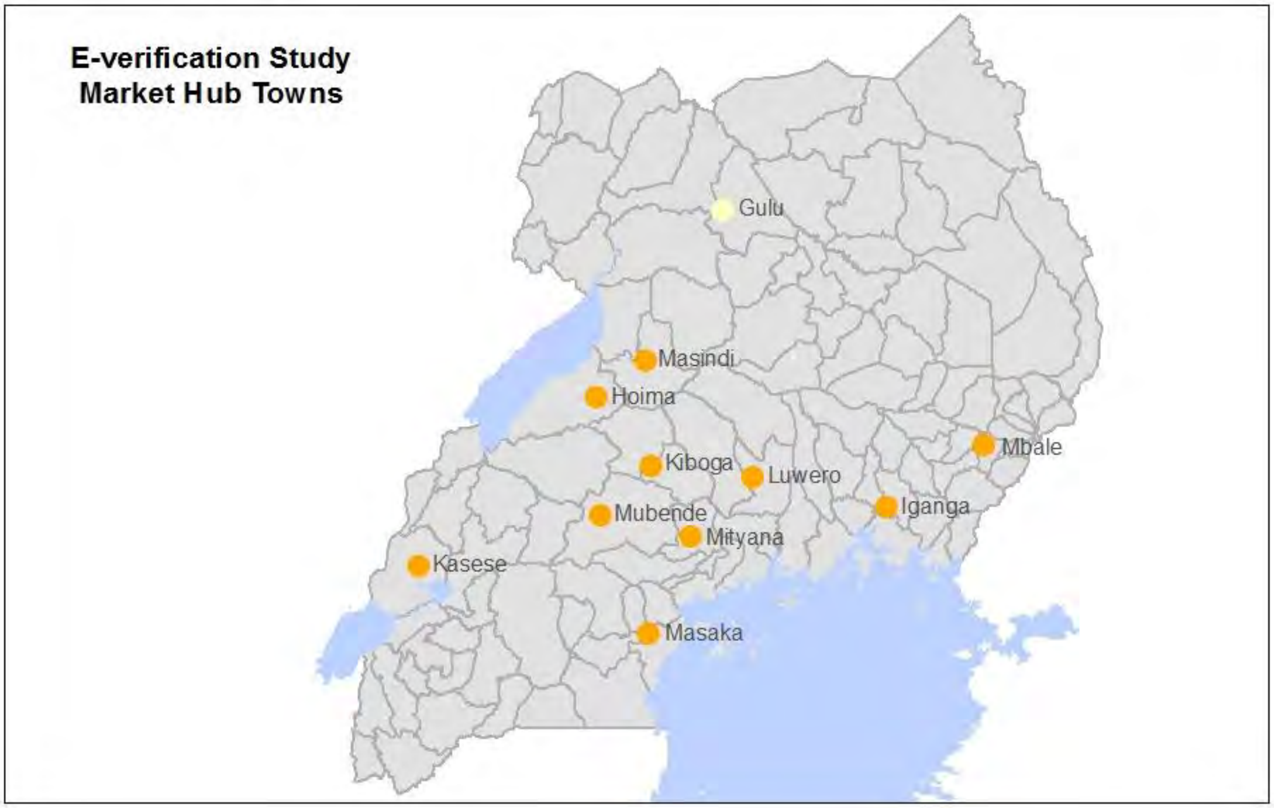
Figure 4.1: Market hubs, market locations, and villages



Ten market hubs were identified alongside Tetra Tech for inclusion in a market survey. These are: Hoima, Iganga, Kasese, Kiboga, Luwero, Masaka, Masindi, Mbale, Mityana, and Mubende (see Figure 4.2).[4] They were selected according to two main criteria: location in a high maize growing area, and probability of receiving e-verified products. The latter criterion is akin to identifying major market hubs where improved agricultural inputs are commonly sold. This list of hubs was identified in part through consultation with staff of the FTF Ag Inputs activity. Hoima and Mityana are not part of the FTF zone of influence. However, including these two districts improves the representativeness of the study area since they are major maize-growing districts.

[4] Gulu was included in the study only for measuring counterfeiting rates as USAID wanted the research team to measure counterfeiting in the Northern region of Uganda. Gulu is not included in any of the analyses presented here.

Figure 4.2 Location of E-verification study market hub towns



A market survey was conducted that gathered lists of the market locations served by these market hubs, of villages served by each market location, and of retail shops in each market location. Using the data gathered during the market survey, 120 market locations were selected for inclusion in the study. During the market hub survey, 139 market locations in the ten hubs were identified. The 120 market locations were selected as the locations with the highest proportion of households in the surrounding villages growing maize.

The market survey also collected data on each village surrounding the market hub, including village population (number of households), proportion of farmers growing maize, and distance to the village from the center of the market hub. From this information, villages were matched into pairs based on degree of similarity in these three criteria. During the market survey, informants were asked to list all villages (LC1s) in the surrounding area from which households would come to source agricultural inputs. Information was collected about each of these villages on: distance to the market location, size of the village, and share of households growing maize. Using this information, villages were paired by similarity. We used a Euclidean distance measure using matrix dissimilarity to choose the pair of LC1s that were most similar. If there was a tie, pairs were drawn at random. The survey team then visited the village pair for each market location, selected another key informant, and updated the information used to pair the villages. They would validate that new information by comparing to the original information. Rules were established to determine when major discrepancies between these variables would warrant dropping the pair or finding another match for the pair. If the original information was not validated, another village would be chosen for validation with the remaining village in the pair, or a new pair would be chosen. This process yielded 120 pairs of villages across the ten market hubs, or 240 LC1s.

Randomized assignment[5] was used to allocate one village in each pair to the encouragement treatment. The other village in each pair serves as the control village, from the standpoint of the encouragement design. Both treatment and control villages would have access to the e-verified products available in their local market location or from any other market products. Further, of the 120 encouragement villages, the same procedure was used to randomly allocate 60 villages to receive price discounts for purchasing EV maize and herbicide. In treatment villages, all households that had a phone received the encouragement treatment in the form of a series of interactive voice recordings (IVR) informing them about the availability of e-verified products in their local market. Encouragement villages were also visited by the survey team who would conduct Community Promotion Meetings (CPM), whereby they would advertise Kakasa products and conduct demonstrations of how to use them. Control villages would have *access* to the e-verified products, but no encouragement messages would be sent to households in those villages, no meetings would be held in these villages, and no discounts would be offered.

4.2 Market location survey and community listing exercise

Once the market hubs and market locations were selected, a market location and community listing exercise was conducted. Market locations were considered to be any area with at least one operating shop or market that sold maize seed, herbicide, or fertilizer. Shops that were near to each other were considered to be in the same market location. From within market locations, a list of all agricultural input retail outlets was obtained from key informants. These shops were then visited for a shop survey.

For each of the 240 LC1s that were selected for inclusion into the study, a community listing exercise (CLE) was conducted, whereby enumerators attempted to visit all households residing in the LC1. Approximately 100 households were interviewed in each of the selected villages. If there were considerably more than 100 households in the village, the team leader drew a village map and consulted village leaders to identify subdivisions in the village that would allow them to select subsections to include in the listing.

4.3 Household Survey

From the community listing exercise, ten households were randomly selected to receive an extensive household survey collecting information on many topics. Seven households who

[5] We used re-randomization as our randomization technique. This involved randomly assigning pairs of villages, testing differences in key variables between treatment and control villages for that particular random assignment, and noting the maximum t-statistic within the set of variables tested. This process was repeated six times (7 total randomizations) and we chose the randomization that had the smallest maximum t-statistic.

reported in the CLE that they owned a mobile phone were selected for inclusion into the household sample, and three households who reported in the CLE that they did not own a mobile phone were also selected for inclusion into the household sample. These households were sampled using simple random sampling.

4.4 Sample size and power calculations

Pre-baseline sample size calculations were performed using data from the HarvestPlus Orange Sweet Potato study, conducted in Mukono, Bukedea, and Kamuli districts in Uganda in 2009. This analysis concluded that a sample of 120 market locations, with 240 villages, each with 10 randomly selected households would enable us to detect reasonably sized treatment effects. Further details can be found in the study's Baseline Report (Ashour et al, 2016).

With these sample sizes, we are able to detect reasonable differences between the treatment (encouragement) and control groups in terms of take-up of both EV products and agricultural inputs in general. For e-verified hybrid maize, we are able to detect differences in take-up of 6.4 percentage points. For e-verified herbicide, due to the high intra-cluster correlation, the sample has the power to detect a difference of 15 percentage points in take-up between encouragement and control. Minimum Detectable Effects (MDE) for the adoption of each input (e-verified or not) are similar; 6.5 percentage points for hybrid maize, 15 percentage points for herbicide, and 4 percentage points for inorganic fertilizer.

5. Surveys

Three rounds of data collection were undertaken: baseline, baseline II, and endline. As explained earlier, a second baseline was conducted to update key variables since the implementation of EV was delayed. The baseline survey was conducted in July-August 2014, the second baseline in January 2016, and the endline in July-August 2017. The endline was conducted two 'growing seasons'[6] after the intervention began and after two seasons of implementation of the encouragement activities.

5.1 Baseline Survey

The baseline data were collected using a series of four survey instruments to gather data at the market level, the village level, and the household level: (1) the market survey was used to identify market locations in the study market hubs, collect shop-level data from all retail shops selling agricultural inputs in each market location, and identify the villages served by each market location; (2) the Community Listing Exercise (CLE), which briefly interviewed all households in the LC1s that were randomly selected for the baseline survey; (3) the household survey was designed to collect information at the household, individual, and plot levels from a sample of ten households in each study village; and (4) the community survey collected information on demographics, land and infrastructure from the LC1 chairperson in each study community.

The market survey is described above. The shop questionnaire consisted of general characteristics of the shop (type of building, etc.), products and sales, demographic information on the shop owner, banking and credit, products sold, their source, their price, and the top brands of maize, herbicide, and fertilizer, and farmer perceptions and recommendations made to farmers.

The CLE yielded interviews on just over 20,000 households. Data were collected on household demographics, farming and input use, purchases of agricultural inputs, and access to and use of mobile phones. Ten farmers in each village also received the household survey instrument. The complete list of modules included in the household survey instrument is provided in the baseline report, but data collected included: a household roster to collect basic demographic information on all current household members such as age, marital status, relationship to household head, and education. It identified the primary agricultural decision maker (PADM) and the secondary agricultural decisionmaker was also identified (SADM). Information on agricultural production was asked regarding each parcel farmed by the household. A parcel is defined as a continuous piece of land under one ownership status. The module captured basic information about all parcels including size, crop choice for the current and previous agricultural seasons, and input use. For every parcel that had any maize production in the current or previous season, the respondent was

[6] There are two growing seasons in Uganda. The first growing season is the main growing season and occurs over January – March. The second growing season is less important, and fewer households grow fewer crops, and it occurs over June – August.

asked specific information about the maize plots on that parcel including size, other crops planted on the plot other than maize, and maize production for the plot. Of the plots listed with any maize production during the current agricultural season, the program randomly selected one plot as a representative maize plot (RMP). The respondent was asked about production details and sales of all crops from the RMP. At the conclusion of the interview the enumerator measured the area of the RMP using a GPS device. The production module also collected data on input use and perceptions of input quality for all plots during the current and previous agricultural seasons and specific input use on the RMP for the current season. An assets module documented the value of household assets, livestock, and enterprise assets. Other modules used to assess household wealth and vulnerability included a non-food expenditure module, questions on access to and use of credit, and exposure to economic shocks. A food consumption module was included to document all food expenditures and consumption of the household in the seven days prior to the interview. A child food frequency module was used to identify the food groups consumed by all children 6–35 months of age in the household during the day and night prior to the interview. Questions were also included on food consumed away from home and coping strategies during times of food shortage.

Data on individual preferences and beliefs was also collected. These modules included both qualitative and quantitative questions asked of the PADM and SADM to assess a number of characteristics about the individual. The qualitative questions were administered during the main part of the interview and included a series of questions to assess risk aversion and ambiguity aversion, a hypothetical game used to measure respondents' time preferences, and a general assessment of respondents' trust. The remainder of this module was conducted after the conclusion of the main survey in a group format with all available PADM respondents from the village (maximum 10 respondents). The respondents were introduced to a series of games that were designed to quantitatively measure risk aversion and ambiguity aversion based on the Holt and Laury (2002) lottery experiments. Each game presented a hypothetical agricultural scenario and respondents were asked to make a choice that would identify their risk or ambiguity preferences. The scenario was explained to the group of respondents and then the respondents selected their choice individually in private. Respondents were also asked about their beliefs of the quality of different agricultural inputs on the market using props to form a distribution to represent their confidence in the quality of the products. For further details on the post-interview games, see Ashour et al. (2018).

The community questionnaire was administered at the time of the household questionnaire by the team leaders who interviewed the LC1 chairperson of each village. Topics covered in the community questionnaire included demographics, land, infrastructure, farming activities, and major events experienced by the community, both positive and negative, in the past two years.

5.2 Second Baseline

The second baseline updated information on certain key variables that were needed, both to understand adoption patterns as well as to update phone numbers since this was the primary means of providing the encouragement intervention. All households were asked demographic questions about the household head and PADM (gender, education, etc.), basic questions on farming (when they last cultivated and what they grew), perceptions of input quality and counterfeiting, risk and ambiguity aversion (qualitative questions), and phone numbers. The ten household survey households were additionally asked questions about input use and purchase of hybrid maize, glyphosate herbicide, and inorganic fertilizer.

A short shop questionnaire was also administered to all shops that existed at that time in each of the 120 market locations. This was a shortened version of the baseline shop questionnaire, and also checked whether any shops had closed down and if any new shops had opened. Questions included topics such as products sold (the top 7 brands of each of herbicide, maize, and fertilizer), prices and quantities, source of products sold, and general characteristics (type of shop, type of building, whether the shop has a bank account, etc.).

5.3 Endline Survey

The endline survey followed a similar structure to the baseline survey. A CLE survey was administered to all households that were listed as part of the first two CLE surveys, a household survey was administered to all households who were listed as household survey households in the baseline survey or second CLE, and a shop survey was also conducted.

The endline CLE included all of the questions from the second CLE (with a less detailed phone module) and also included a detailed section on experience with EV products and with the encouragement interventions. General questions were asked first, regarding whether the household had heard of the two EV products (E-tag and Ag-Verify), where they heard about it, whether they purchased any, and whether they verified. Then, for each of the two seasons in which the IVR calls and CPMs were made, respondents were asked whether they had received a call or attended a meeting, whether they thought they were useful, whether they received the discount vouchers, and whether they used them.

The endline household survey also asked the same questions as the baseline household survey (including the post-interview games) with the addition of the detailed module on EV and the encouragement activities. The short shop questionnaire from the second round of data collection was also administered in the shops currently in operation in all 120 market locations.

5.4 Data capture and cleaning

During the survey interviews, enumerators recorded respondents' responses to the survey questions laptop computers using Computer Assisted Personal Interview (CAPI) software called

SurveyBe. A data capture protocol was developed to ensure that data were saved and backed up on a daily basis and reduce the risk of lost data files. Team leaders used a flash drive to collect data from each enumerator's computer on a daily basis. All interview files were saved on the team leaders' computers and also uploaded to a shared, password protected and encrypted Dropbox folder. The IFPRI team was able to access the uploaded files and check the data. In addition, the team leaders conducted random checks of the household interviews each evening using a flagging protocol to identify data errors.

Once the surveys were completed, the data were labeled and cleaned. All identifying information, such as names, phone numbers, and locations of respondent households, was removed from the datasets for privacy purposes before the data were made publicly available.

6. Estimating Treatment Effects

6.1 Identification strategy

Individual and household characteristics, coverage of other interventions, and community characteristics should be roughly equal across both the encouragement and control treatment arms and within and between the discount and no discount arms and the discount arms compared to the control group as a result of randomization. Randomization should limit the effects of confounding variables on the impact estimates. Thus, the control group can be taken as a valid proxy for the counterfactual situation of the other study arms in the absence of the intervention. Consequently, average differences in outcomes between the treatment arms after intervention can be interpreted as causal impacts; that is, they can be interpreted as being caused by, rather than simply correlated with, the interventions.

Encouragement designs provide an effective method for measuring the causal impact of a program that is available to all households within a study area when take-up of the program is not universal. In an encouragement design, households living in communities surrounding markets with e-verified products were sent IVR messages informing the households that these products are available and encouraging them to purchase the products, as well as community promotion meetings whereby demonstrations and advertisements are made. The encouragement design, whereby some people receive a "nudge" to purchase these products, allows a clean test of whether the use of e-verified products affects key household-level outcomes, such as adoption of improved inputs and yields. The effect of access to e-verified products is identified off of the difference in take-up rates of e-verified inputs between encouraged and non-encouraged communities created by the encouragement IVR messages and community meetings.

The estimation methodology measures the effect of the e-verification program on use of high-quality inputs and other outcomes of interest by estimating the impact of purchasing any e-verified products on these outcomes, based on experimentally induced differences in predicted e-verification purchasing behavior between treatment households exposed to the encouragement messages and control households who did not receive the encouragement promotion messages. All outcomes are measured at endline, controlling for the value of the outcome in the two previous survey rounds, baseline I, baseline II. The first baseline survey was carried out in July-August 2014. The second baseline was carried out in January-February 2016 and was necessitated because the roll out of e-verified inputs was delayed. By the time the program was ready to roll out, the 2014 baseline was a bit outdated, particularly in terms of phone numbers (which were a central component by which to implement our encouragement design) and thus the research team conducted a shorter second baseline survey to update key variables. The endline survey was carried out in July-August 2017.

The primary outcomes for research questions 1 and 2 include:

1. The rate of take-up of improved agricultural inputs (glyphosate herbicide, inorganic fertilizer, and hybrid maize seeds)
2. The increase in farm yields, incomes, and net incomes

The impact of EV on these outcomes will be measured in the same way, using a standard empirical model for encouragement designs, described below.

6.1.1 Household-level outcomes

Consider an empirical model of the impact of e-verification on farm income. We want to estimate a model of farm income as a function of the rate of adoption of e-verified hybrid maize. However, the roll out of e-verification will be non-random across the study areas because access to EV products may be higher in areas closer to Kampala or wherever distributors have easier access to retail shops. Similarly, households that choose to adopt EV products may be different from other households in the sense of being more willing to risk trying a new product. These location- and household-level differences that are not random lead to biased estimates of the impact of e-verification on farm income. We will use the encouragement, which will randomly induce differences in take-up rates of e-verified products between encouraged and non-encouraged villages, to identify the effects of e-verification. Since the encouragement is randomly allocated across villages, this will allow us to estimate the causal effect of e-verification on the outcomes of interest, without any confounding variables that would result in biased impact estimates.

We begin the analysis by estimating the direct effect of exposure to the encouragement intervention on the primary outcomes. We estimate these effects in a "reduced form" model, which measures the "Intent to Treat (ITT)" estimate of the experimental encouragement intervention on the study population. This will be the average *effect of the encouragement interventions* (voice messages and community meetings) on adoption of high-quality inputs (hybrid maize, herbicide, and fertilizer) as well as farm level outcomes such as yields, farm income, and profits. The estimating equation will be as follows:

$$y_{ijpt} = \pi_o + \pi_1 E_j + \pi_2 P_{ijt0} + \boldsymbol{\pi_3 y_{ijpt0}} + \gamma_m + \varepsilon_{ijpt} \qquad (1)$$

where y_{ijpt} is the outcome variable for individual *i* in village *j* of phone type (phone or no phone) *p*, in time *t* (the endline). E_j is a dummy variable for the individual's village having randomly been assigned as an encouragement village, P_{ijt0} is a dummy variable for having a phone at baseline[7], $\boldsymbol{y_{ijpt0}}$ is a vector containing the baseline outcome variable[8] (we use the two seasons covered in the

[7] We include this because our sample was drawn such that out of the ten randomly sampled households in the village, 7 had phones at baseline and 3 did not. These three households are likely to be systematically different from phone owning households and so controlling for this variable will reduce bias on the coefficient of interest, π_1.

baseline survey, second season 2013 and first season 2014),[9] γ_m is a set of strata fixed effects (120 market location dummy variables), and ε_{ijpt} is the error term. The error term is clustered at the level of the village to allow for unobserved variables to be correlated with each other within the village. This equation provides an indication of the effectiveness of the encouragement intervention, by estimating whether households in villages randomly assigned to receive the encouragement messages are more likely to have used high-quality inputs and increased farm productivity and profits.

The reduced form model of encouragement effects in (1) controls for phone ownership. Since a significant component of the encouragement intervention was delivered as phone messages, those who do not own phones could not benefit as much from the intervention (apart from the community meetings – but even information on those was often spread via phone). However, non-phone owners may have heard about EV products from phone owning households and other members and leaders of their community, through shop owners, or other advertising. Also, both households that own phones and those that do not could have heard about the e-verification program through the community information meetings run as part of the encouragement intervention. Some households who do not own a mobile phone have been included in the sample (three per LC1, on average). In addition, some members of the community who do own a mobile phone will not receive encouragement messages, some households may not purchase e-verified products, or even if they do, they may not authenticate them. Thus, ensuring to account for households who do not own phones is very important in understanding adoption behavior, as not accounting for them could over or understate the effect of the program. Regressions are weighted to account for this sampling procedure: households with phones are weighted by 1/7 to account for the likelihood of their inclusion in the household sample being 7 out of 10, and households without phones are weighted by 1/3 to account for the likelihood of their inclusion in the household sample being 3 out of 10.

We also estimate the equation above including several baseline household level controls as a robustness check (reported in Appendix A). If there are not statistically significant differences across encouragement and control households in baseline characteristics, the coefficient estimates should not be affected. However, if there are differences in observed variables (or even in unobserved variables that are correlated with included observables) including baseline values of these variables in the regression will improve precision in our estimates. We include the following baseline control variables in the specification: household head's age, whether female, literacy status, and whether they finished primary school, as well as these same variables for the primary agricultural decision maker (PADM). We also include household size, total land area owned, and the number of parcels owned. As a way of improving our power to detect treatment effects by maintaining a larger sample size, we also create a dummy variable for households who have

[9] For some outcomes, we have only one baseline value (for example profits, which can only be calculated in the household survey at baseline and endline). For other outcomes, such as use of inputs, we have data by season and can often include two baseline seasons as control variables in the regression to increase precision (McKenzie, 2012).

missing baseline data and replace their missing baseline data with the average value for that variable and include the dummy variable for missing baseline data in the regression.

The main impact of interest is the *effect of e-verified products on input adoption and farm outcomes*. We measure the effect of the e-verification products themselves (as opposed to measuring the impact of only the encouragement activities) on outcomes using an "instrumental variables" approach. This approach estimates the "Local Average Treatment Effect (LATE)," which is the impact for those households who are induced to take up the e-verified product due to the encouragement campaign (not for those households who would have taken it up anyway or for those who would never take it up). This approach includes two stages (two regressions). First, we estimate the effect of the encouragement interventions on awareness of e-verified inputs and obtain estimates of the predicted probability of being aware of e-verified products. The encouragement interventions are implemented to cause higher awareness in encouragement communities than control communities in a randomly assigned way. Then, in the second stage, we estimate the effect of these predicted probabilities of being aware of EV from the first stage regression on outcomes such as adoption and farm incomes.

This two-stage model of the randomized encouragement design is needed to obtain unbiased estimates of the impact of e-verification because, without the encouragement campaign, awareness of e-verified products is not random. Households that are aware of e-verified products or have tried them on their own are likely to have different characteristics than households on average, such as being more innovative or having better knowledge of new technologies. Comparing outcomes for these "innovative" households to those who are not aware of e-verified products or have not tried them would lead to biased impact estimates as the difference in outcomes would include the effect of households using e-verification being more innovative. A common approach to designing a study to remove the effect of these characteristics is to conduct a randomized control trial, where e-verified products are made available only in randomly selected markets. However, it was not possible to control which markets got access to e-verified products (all households in the study did). The randomized encouragement design is a good alternative study design in this case because the randomly assigned encouragement campaign creates experimental variation in awareness or exposure to the e-verified products. This should lead to higher take-up of these products in randomly assigned encouragement communities than in randomly-assigned control communities around the same markets. We expect to observe awareness or adoption of e-verified products in both encouragement treatment and control communities, but higher adoption in encouragement treatment communities. The random variation comes from the first stage equation – the encouragement to be more aware. If we estimate the second step without estimating the first step, our impact estimates would be biased.

The regression equations are as follows:

$$EV_{ijpt} = \alpha_0 + \alpha_1 E_j + \alpha_2 P_{ijt0} + \boldsymbol{\alpha_3 y_{ijpt0}} + \gamma_m + \epsilon_{ijpt} \tag{2}$$

$$y_{ijpt} = \beta_0 + \beta_1 \widehat{EV_{ijpt}} + \beta_2 P_{ijt0} + \boldsymbol{\beta_3 y_{ijpt0}} + \gamma_m + \varepsilon_{ijpt}. \tag{3}$$

Where EV_{ijpt} is whether the household has heard of either e-tag or Ag-verify, $\widehat{EV_{ijpt}}$ are the predicted values of awareness of tagged products for individual i at time t from equation (2), and other variables are defined as before. In equation (2), α_2 will measure the impact of the encouragement on knowledge of the product assurance schemes, and in equation (3), β_1 will measure the effect of e-verification on adoption of high-quality inputs, yields, incomes, and profits. In this way, the random variation in the encouragement is used to identify differences in take-up, which are then used to identify the effect of take-up and farm incomes.

The reason that the endogenous variable is whether the respondent has heard of E-tag/Ag-verify is that the encouragement campaign was supposed to increase awareness of the tagged products. Then, households were supposed to take up verified products, but the encouragement design was supposed to first improve knowledge. Just because a household did not purchase tagged products in the two seasons for which the endline survey was conducted, does not mean that a household would not purchase them in the future since they know about them. It may take time for some households to see the experiences of others before they adopt. Further, adoption of agricultural inputs in general are often the outcome measures of interest. The adoption of E-tag/Ag-verify variables are highly correlated with these outcomes and are actually a sub-set of the adoption that occurs. Thus, there is a mechanical relationship between the two. For these reasons, we feel that the correct measure to estimate the impacts of the verification intervention is whether a household has heard of the E-tag/Ag-verify programs. If we instead do use purchase of any E-tag/Ag-verify product in the first stage, the results are quite similar, and these results are reported in Appendix B.

This two-stage estimation method of estimation removes any bias in the estimate of the impact of e-verification on farm incomes due to nonrandom availability of e-verification or selection in take-up of e-verification by farmers. Regressions are also weighted as above, baseline missing values replaced as above, and standard errors are clustered at the level of the village. A separate set of tables that include the same baseline household level controls as listed above is also estimated as a robustness check in Appendix A.

6.1.2 Sample Size and Attrition

At baseline, the goal was to interview 2,400 households on the long baseline household survey in 2014 (ten households in each of 240 villages). We were able to successfully interview 2,375 households at endline in 2017 (some households on the sample lists refused or were not available and could not be replaced due to time constraints). Of these 2,375 households, 1,996 were interviewed at endline. There were 51 households who were interviewed at endline in 2017 and who were part of the original targeted sample at baseline but were not interviewed at baseline. We include these households in the analysis of endline outcomes (by replacing their missing baseline data with the mean for a particular variable and creating a dummy variable equal to one if baseline data is missing for that household), and thus our final analysis sample comprises 2,047 households. This implies a rate of attrition of 14 percent.

We conduct analyses to ascertain whether attrition from the sample differs in encouragement versus control communities, and within encouragement communities whether attrition differs between discount and non-discount communities. Table 6.1 shows a regression of the outcome of a household having attrited between baseline and the second baseline on a dummy variable for the encouragement assignment only in Panel A and on both the encouragement and discount assignment in the first column in Panel B. In the second column, we show the same for households who were interviewed at baseline but not at endline. In the third column we report attrition from the second baseline to the endline. In columns (4) and (5) we report attrition from the first to second baseline and from the baseline to endline, respectively, excluding villages that refused to be interviewed. We find that between the original baseline and the second baseline, as well as between the second baseline and the endline, there was no differential attrition across the encouragement and the discount treatment groups compared to the control group. Between baseline and endline, we do not see differential attrition for the discount group, but we do see a bit of a difference for encouragement households, who were more likely not to be tracked at endline. However, this difference is quite small; households in encouragement villages are 2.7 percentage points less likely to be interviewed at endline compared to control households. The baseline and endline were three years apart and thus it would be more difficult to track households. We also note that at endline, there was one entire village that refused to be interviewed.

Table 6.1: Attrition across 3 rounds

	(1) Baselined HHs NOT interviewed at CLE2	(2) Baselined HHs NOT interviewed at endline	(3) CLE2 HHs NOT interviewed at endline	(4) Baseline HHs NOT interviewed at CLE2, excluding villages that refused	(5) Baseline HHs NOT interviewed at endline, excluding villages that refused
Panel A: Encouragement group					
Encouragement	0.007	-0.028**	-0.018	0.000	-0.032***
	(0.013)	(0.011)	(0.011)	(0.013)	(0.011)
Observations	2374	2374	1894	2344	2364
R^2	0.134	0.087	0.085	0.094	0.067
Control group mean	0.199	0.173	0.124	0.199	0.173
Panel B: Encouragement and Discount groups					
Encouragement	0.001	-0.026**	-0.020*	0.001	-0.026**
	(0.017)	(0.013)	(0.012)	(0.018)	(0.013)
Discount	0.013	-0.003	0.006	-0.001	-0.011
	(0.027)	(0.022)	(0.022)	(0.025)	(0.022)

Observations	2374	2374	1894	2344	2364
R^2	0.134	0.087	0.085	0.094	0.068
Control group mean	0.199	0.173	0.124	0.192	0.173

Note: Standard errors are reported in parentheses. Regressions control for market location fixed effects to account for stratification and standard errors clustered at the village level, and also weight observations according to their probability of selection into the sample based on whether the household owned a phone in the first CLE. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

7. Impact Estimates of the Effect of the E-Verification Project

In this section, we will present the causal impacts of the e-verification project on a number of outcomes, including take up of e-verified products, take up of productivity-enhancing agricultural inputs in general, yields, profits, and wellbeing. We will begin by presenting some summary statistics and checking for baseline balance across treatment groups as well as summary statistics on experience with e-verification and with our interventions, followed by the reduced form regressions, and finally, the IV regression estimates.

7.1 Summary Statistics and Baseline Balance

Table 7.1a displays summary statistics on baseline demographic characteristics of households. First, means are reported for the control group (C), the encouragement group (E), and the discount group (D) in columns (1) to (3). Next, p-values for whether there are statistically significant differences in means between groups are reported. We compare the encouragement group to the control group, the encouragement group to the discount group, and the encouragement and discount group (to check whether the full intervention is balanced) to the control group in columns (4) to (6).[10] P-values below 0.01 indicate a less than one percent probability that the means are equal, p-values between 0.01 and 0.05 indicate a 1 to 5 percent probability that the means are equal, and p-values between 0.05 and 0.1 indicate a 5 to 10 percent probability that the means are equal. Finally, in column (7) we report the number of observations in the data for that particular variable.

On average, the head of the household is approximately 46 years old, about three quarters are male, and about 37 percent have completed primary school. Approximately 65 percent of household heads are literate. The average age of the primary agricultural decision maker is approximately 45 years old, approximately 74 percent are male, and about 37 percent have completed primary

[10] Means and regressions also weight observations according to their probability of selection into the sample based on whether the household owned a phone in the first CLE. P-values for differences in means are derived from two separate regressions. The first is a regression of the variable of interest on a dummy variable for being assigned to the encouragement group without including the discount group and comparing testing whether the coefficient on the encouragement dummy is significantly different from zero. The second is a regression of the outcome of interest on dummy variables for the encouragement treatment and the discount treatment, with tests for the equality of the encouragement and discount groups, as well as a test for whether the coefficient on the dummy variable for belonging to the discount group is significantly different from zero. All regressions include a dummy variable for not owning a phone and control for market location fixed effects to account for stratification and standard errors clustered at the village level.

school, while approximately 66 percent are literate. This suggests that the household head and primary agricultural decision maker are often the same person, but that sometimes the agricultural decision maker is the spouse or female head of household. However, their education, age, and literacy are very similar. Households across treatment groups have approximately 5.6 members on average. We calculate several asset indices, for household durables, land, livestock, enterprise assets, and finally all of the assets listed above. We use principal components analysis and report the first principal component in each case.

We find that households across treatment arms look very similar to one another. We have very good balance on household demographic characteristics. Slightly more households in the control group have female primary agricultural decision makers, and the overall asset index is lower in the control group compared to the other groups (however the sub-indices are well balanced). Overall, there are no statistically significant differences between any of the household demographics between any of the groups.

Table 7.1a: Baseline Household Demographic Characteristics and Asset Indices

	Mean (sd)			Difference in means (p-value)			
	Control (C) (1)	Encouragement (E) (2)	Discount (D) (3)	(E)-(C) (4)	(D)-(E) (5)	(E)+(D)-(C) (6)	N (7)
Household head – Age	46.836 (16.093)	46.236 (16.261)	46.604 (17.217)	0.373	0.551	0.765	1,978
Household head – Male	0.742 (0.437)	0.735 (0.442)	0.740 (0.439)	0.685	0.763	0.946	1,978
Household head - Completed primary	0.367 (0.482)	0.380 (0.486)	0.374 (0.484)	0.346	0.730	0.586	1,978
Household head - Literate	0.663 (0.473)	0.661 (0.474)	0.667 (0.472)	0.804	0.682	0.435	1,978
Decision maker - Age	45.695 (15.819)	44.900 (15.770)	45.348 (16.581)	0.151	0.447	0.321	1,994
Decision maker - Male	0.661 (0.474)	0.639 (0.485)	0.649 (0.486)	0.193	0.600	0.536	1,994
Decision maker - Completed primary	0.206 (0.405)	0.214 (0.410)	0.229 (0.421)	0.459	0.317	0.123	1,992
Decision maker - Literate	0.648 (0.478)	0.654 (0.476)	0.657 (0.475)	0.911	0.842	0.393	1,992
Number of household members	5.556 (2.948)	5.623 (2.892)	5.649 (3.019)	0.466	0.819	0.422	2,041
Durables asset index	-0.001 (1.507)	0.057 (1.435)	0.062 (1.488)	0.271	0.916	0.502	1,960
Land asset index	0.019 (1.051)	-0.026 (0.985)	0.008 (1.018)	0.299	0.366	0.987	1,986
Livestock asset index	-0.003 (1.365)	0.024 (1.374)	0.067 (1.499)	0.431	0.381	0.506	1,969
Enterprise asset index	-0.027 (1.050)	0.044 (1.245)	0.073 (1.285)	0.144	0.556	0.231	1,968
All asset index (except land asset)	-0.010 (1.768)	0.066 (1.692)	0.097 (1.767)	0.227	0.566	0.950	1,948

Notes: Standard deviations are reported in parentheses. P-values for differences in means are derived from two separate regressions. The first is a regression of the variable of interest on a dummy variable for being assigned to the encouragement group without including the discount group and comparing testing whether the coefficient on the encouragement dummy is significantly different from zero. The second is a regression of the outcome of interest on dummy variables for the encouragement treatment and the discount treatment, with tests for the equality of the encouragement and discount groups, as well as a test for whether the coefficient on the dummy variable for belonging to the discount group is significantly different from zero. All regressions include a dummy variable for not owning a phone and control for market location fixed effects to account for stratification and standard errors clustered at the village level. Means and regressions also weight observations according to their probability of selection into the sample based on whether the household owned a phone in the first CLE. Asset indices were constructed using principal components analysis and reporting the first principal component. Stars denote statistical significance as follows: * $p < 0.1$, ** $p < .05$, *** $p < 0.01$.

In Table 7.1b we present baseline summary statistics and tests for balance on variables relating to farm characteristics and adoption of productivity enhancing inputs. The table can be interpreted in the same way as Table 7.1a. The baseline survey covered two growing seasons: the second season of 2013 and first season of 2014. The first season is generally the main maize growing season (and main agricultural season in general). The table is structured in the same way as Table 7.1a. On average, households own two parcels of land and own approximately 5-6 acres of land, with control group households owning slightly fewer parcels and owning less land. Households grow an average of five crops during the first season of 2014 and about four different crops during second season of 2013. Use of hybrid maize seed is relatively low with only about ten and eight percent of the sample using hybrid maize in first season 2014 and second season 2013, respectively. However, we do see that almost all households in the sample (an average of 93 percent) are growing maize in first season 2014 and approximately 93 percent of the sample is growing maize in the second season of 2013. Similar patterns are observed in the rest of the variables as well. Approximately 34 percent of the sample uses glyphosate herbicide in the first season and approximately 26 percent in the second season. Use of any herbicide is slightly higher, but almost all the herbicide used in Uganda is glyphosate. Fewer households are using fertilizer, with approximately ten percent of households using inorganic fertilizer and 15 percent using any fertilizer in the first season of 2014 and approximately 5 percent using inorganic fertilizer and 9 percent using any fertilizer in the second season of 2013.

Once again, our baseline characteristics regarding farming and input use are fairly well balanced. There are slight differences between groups in terms of input usage, however, with 48 tests involved what we find is what we should expect statistically by chance.

Table 7.1b: Baseline Household Farm Characteristics and Input Use

	Mean (sd)			Difference in means (p-value)			
	Control (C)	Encouragement (E)	Discount (D)	(E)-(C)	(D)-(E)	(E)+(D)-(C)	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Number of parcels owned	1.993	2.070	2.104	0.009	0.439	0.020	2,041
	(1.034)	(1.094)	(1.152)				
Total land area, acres	6.125	5.157	5.237	0.081	0.866	0.322	1,991
	(16.271)	(11.481)	(11.599)				
Number of crops grown in FS14	4.323	4.401	4.461	0.080	0.368	0.012	1,990
	(1.416)	(1.402)	(1.416)				
Used hybrid maize seeds in FS14	0.096	0.096	0.102	0.931	0.605	0.683	1,973
	(0.294)	(0.295)	(0.303)				
Used any maize seeds in FS14	0.931	0.938	0.933	0.328	0.567	0.365	1,985
	(0.254)	(0.240)	(0.250)				
Used glyphosate herbicide in FS14	0.320	0.352	0.379	0.109	0.296	0.112	1,961
	(0.467)	(0.478)	(0.486)				
Used any herbicide in FS14	0.344	0.380	0.408	0.074	0.311	0.148	1,981
	(0.475)	(0.486)	(0.492)				
Used inorganic fertilizer in FS14	0.097	0.103	0.116	0.633	0.247	0.125	1,980
	(0.296)	(0.304)	(0.321)				
Used any fertilizer in FS14	0.150	0.150	0.164	0.969	0.379	0.095	1,980
	(0.357)	(0.357)	(0.370)				
Number of crops grown in SS13	3.640	3.764	3.761	0.070	0.959	0.066	1,990
	(1.692)	(1.584)	(1.576)				
Used hybrid maize seeds in SS13	0.080	0.066	0.074	0.144	0.390	0.842	1,900
	(0.272)	(0.248)	(0.262)				
Used any maize seeds in SS13	0.825	0.825	0.808	0.966	0.345	0.414	1,910
	(0.380)	(0.380)	(0.394)				
Used glyphosate herbicide in SS13	0.257	0.260	0.287	0.696	0.317	0.357	1,886
	(0.437)	(0.439)	(0.453)				
Used any herbicide in SS13	0.272	0.290	0.315	0.476	0.353	0.058	1,910
	(0.445)	(0.454)	(0.465)				
Used inorganic fertilizer in SS13	0.057	0.051	0.051	0.320	0.978	0.702	1,905
	(0.233)	(0.220)	(0.220)				
Used any fertilizer in SS13	0.092	0.090	0.085	0.724	0.691	0.758	1,908
	(0.290)	(0.286)	(0.279)				

Notes: Standard deviations are reported in parentheses. P-values for differences in means are derived from two separate regressions. The first is a regression of the variable of interest on a dummy variable for being assigned to the encouragement group without including the discount group and comparing testing whether the coefficient on the encouragement dummy is significantly different from zero. The second is a regression of the outcome of interest on dummy variables for the encouragement treatment and the discount treatment, with tests for the equality of the encouragement and discount groups, as well as a test for whether the coefficient on the dummy variable for belonging to the discount group is significantly different from zero. All regressions include a dummy variable for not owning a phone and control for market location fixed effects to account for stratification and standard errors clustered at the village level. Means and regressions also weight observations according to their probability of selection into the sample based on whether the household owned a phone in the first CLE. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

7.2 Experience with EV and with the Encouragement

In this section we will describe the households' experiences with the EV products and with our encouragement interventions. The tables in this section follow the same format as the tables in section 7.1 and can be interpreted in the same way, except we report on endline variables rather than baseline variables. Thus, P-values below 0.01 indicate a less than one percent probability that the means are equal, p-values between 0.01 and 0.05 indicate a 1 to 5 percent probability that the means are equal, and p-values between 0.05 and 0.1 indicate a 5 to 10 percent probability that the means are equal. Table 7.2 reports on households' experiences with the E-tag labeled herbicide and maize, and with the Ag-verify labeled maize.

Table 7.2: Experience with E-tag and Ag-verify products - Endline

	Mean (sd)			Difference in means (p-value)			
	Control (C) (1)	Encouragement (E) (2)	Discount (D) (3)	(E)-(C) (4)	(D)-(E) (5)	(E)+(D)-(C) (6)	N (7)
Experience with E-tag							
Recognized E-tag label	0.020	0.490	0.504	0.000	0.500	0.000	2,060
	(0.141)	(0.500)	(0.500)				
Heard of E-tag	0.111	0.667	0.676	0.000	0.682	0.000	2,060
	(0.314)	(0.471)	(0.469)				
Heard about E-tag via network	0.016	0.090	0.098	0.000	0.436	0.000	2,060
	(0.124)	(0.286)	(0.297)				
Heard about E-tag via media	0.087	0.149	0.156	0.000	0.626	0.000	2,060
	(0.281)	(0.357)	(0.363)				
Heard about E-tag via shopkeeper	0.002	0.011	0.008	0.001	0.387	0.113	2,060
	(0.045)	(0.105)	(0.089)				
Heard about E-tag via community meetings	0.005	0.528	0.539	0.000	0.621	0.000	2,060
	(0.072)	(0.499)	(0.499)				
Heard about E-tag via phone messages	0.006	0.155	0.178	0.000	0.105	0.000	2,060
	(0.078)	(0.362)	(0.383)				
First heard in FS17	0.063	0.201	0.211	0.000	0.414	0.000	2,060
	(0.244)	(0.401)	(0.409)				
First heard in SS16	0.038	0.393	0.389	0.000	0.817	0.000	2,060
	(0.191)	(0.489)	(0.488)				
First heard before SS16	0.010	0.073	0.078	0.000	0.635	0.000	2,060
	(0.101)	(0.260)	(0.269)				
Experience with Ag-verify							
Recognized Ag-verify label	0.009	0.194	0.213	0.000	0.204	0.000	2,060
	(0.096)	(0.396)	(0.410)				
Heard of Ag-verify	0.017	0.183	0.199	0.000	0.251	0.000	2,060
	(0.128)	(0.387)	(0.400)				
Heard about Ag-verify via network	0.002	0.010	0.010	0.009	0.955	0.288	2,060
	(0.045)	(0.099)	(0.100)				
Heard about Ag-verify via media	0.012	0.035	0.036	0.000	0.826	0.001	2,060
	(0.111)	(0.184)	(0.187)				
Heard about Ag-verify via community meetings	0.002	0.167	0.185	0.000	0.160	0.000	2,060
	(0.045)	(0.373)	(0.389)				
Heard about Ag-verify via phone messages	0.000	0.026	0.031	0.000	0.396	0.000	2,060
	(0.000)	(0.159)	(0.174)				
First heard in FS17	0.006	0.116	0.122	0.000	0.628	0.000	2,060
	(0.079)	(0.321)	(0.328)				
First heard in SS16	0.010	0.093	0.105	0.000	0.297	0.000	2,060
	(0.101)	(0.291)	(0.306)				
First heard before SS16	0.001	0.005	0.010	0.024	0.020	0.002	2,060
	(0.032)	(0.070)	(0.100)				

Notes: Standard deviations are reported in parentheses. P-values for differences in means are derived from two separate regressions. The first is a regression of the variable of interest on a dummy variable for being assigned to the encouragement group without including the discount group and comparing testing whether the coefficient on the encouragement dummy is significantly different from zero. The second is a regression of the outcome of interest on dummy variables for the encouragement treatment and the discount treatment, with tests for the equality of the encouragement and discount groups, as well as a test for whether the coefficient on the dummy variable for belonging to the discount group is significantly different from zero. All regressions include a dummy variable for not owning a phone and control for market location fixed effects to account for stratification and standard errors clustered at the village level. Means and regressions also weight observations according to their probability of selection into the sample based on whether the household owned a phone in the first CLE. Stars denote statistical significance as follows: * $p < 0.1$, ** $p < .05$, *** $p < 0.01$.

We see that very few people in the control group recognize the E-tag label, and that only 11 percent had heard about E-tag. In the encouragement and discount groups, however, almost half of households recognized the label and two thirds had heard of E-tag. Most people had heard about E-tag through community meetings (these are not necessarily the community promotions that we had organized as part of the encouragement; AgInputs and E-tag also had road shows), and most people first heard about E-tag in the first season of 2016 when the products were first launched.

Once again, very few people in the control group recognized the Ag-verify label or had heard about Ag-verify, but about 20 percent recognized the label and had heard about it in the encouragement and discount groups. Most people had heard about Ag-verify through phone messages. There were other phone campaigns occurring at the time, but it is likely that most of this is driven by the phone messages that we sent out as part of the encouragement. Most people had also first heard of Ag-verify in the first season of 2016. In general, knowledge of Ag-verify was lower, but this was expected since it was a much smaller campaign.

When we examine purchasing decisions, about 20 percent of households in the encouragement groups had purchased any E-tag products and had purchased E-tag herbicide specifically in first season 2017, but very few had purchased E-tag or any herbicide in second season 2016. This is likely because farmers tend to purchase their inputs in the first growing season of the year and then continue to use them (if they are still available during the second growing season). It appears that very few people purchased E-tag maize at all, except for in the discount group during first season of 2017. Less than one percent of households in the control group had purchased any E-tag maize seeds, just over one percent of households in the encouragement group had purchased E-tag maize seeds, but four percent of households in the discount group purchased E-tag maize seeds in first season 2017. Again, this is the main season during which households tend to purchase high quality inputs.

Table 7.3 follows the same format as Table 7.2 and examines the experiences that households had with our encouragement interventions: the phone messages, the community meetings, and the discounts. As a reminder, the interactive phone messages and community meetings were held in both second season of 2016 and first season of 2017. The discount vouchers were only distributed during the first season of 2017. The data in this table are recall data from the endline household survey.

Table 7.3: Exposure to Treatment Interventions at Endline

	Mean (sd)			Difference in means (p-value)			
	Control (C) (1)	Encouragement (E) (2)	Discount (D) (3)	(E)-(C) (4)	(D)-(E) (5)	(E)+(D)-(C) (6)	N (7)
First season 2017							
Attended community meeting	0.002	0.405	0.424	0.000	0.386	0.000	2,048
	(0.045)	(0.491)	(0.495)				
Found meeting helpful	0.002	0.396	0.418	0.000	0.291	0.000	2,048
	(0.045)	(0.489)	(0.494)				
Received voice messages	0.010	0.284	0.312	0.000	0.077	0.000	2,048
	(0.101)	(0.451)	(0.464)				
Interacted with voice messages	0.004	0.190	0.204	0.000	0.321	0.000	2,048
	(0.064)	(0.393)	(0.403)				
Tried to call back	0.002	0.086	0.101	0.000	0.093	0.000	2,048
	(0.045)	(0.281)	(0.302)				
Thought the calls were too few	0.002	0.056	0.061	0.000	0.487	0.000	2,048
	(0.045)	(0.229)	(0.239)				
Thought the calls were just right	0.004	0.182	0.204	0.000	0.110	0.000	2,048
	(0.063)	(0.386)	(0.403)				
Thought the calls were too many	0.004	0.046	0.048	0.000	0.831	0.000	2,048
	(0.064)	(0.210)	(0.214)				
Found calls useful	0.006	0.253	0.272	0.000	0.220	0.000	2,048
	(0.078)	(0.435)	(0.446)				
Received maize coupon	0.001	0.164	0.299	0.000	0.000	0.000	2,048
	(0.032)	(0.370)	(0.458)				
Received herbicide coupon	0.001	0.162	0.295	0.000	0.000	0.000	2,048
	(0.032)	(0.369)	(0.457)				
Redeemed maize coupon	0.000	0.020	0.038	0.000	0.006	0.000	2,048
	(0.000)	(0.142)	(0.190)				
Redeemed herbicide coupon	0.001	0.051	0.094	0.000	0.000	0.000	2,048
	(0.032)	(0.221)	(0.292)				
Attempted to authenticate E-tag maize	0.001	0.015	0.024	0.000	0.027	0.000	2,048
	(0.032)	(0.121)	(0.153)				
Successfully authenticated E-tag maize	0.001	0.014	0.022	0.000	0.031	0.000	2,036
	(0.032)	(0.117)	(0.148)				
Attempted to authenticate E-tag herbicide	0.005	0.121	0.146	0.000	0.054	0.000	2,048
	(0.072)	(0.326)	(0.354)				
Successfully authenticated E-tag herbicide	0.005	0.115	0.140	0.000	0.067	0.000	1,979
	(0.072)	(0.320)	(0.348)				
Second season 2016							
Attended community meeting	0.005	0.424	0.431	0.000	0.758	0.000	2,048
	(0.072)	(0.494)	(0.496)				
Found meeting helpful	0.004	0.409	0.417	0.000	0.721	0.000	2,048
	(0.064)	(0.492)	(0.493)				
Received voice messages	0.004	0.232	0.233	0.000	0.951	0.000	2,048
	(0.064)	(0.422)	(0.423)				
Interacted with voice messages	0.002	0.156	0.149	0.000	0.596	0.000	2,048
	(0.045)	(0.363)	(0.357)				
Tried to call back	0.001	0.071	0.071	0.000	0.915	0.000	2,048
	(0.032)	(0.256)	(0.258)				
Thought the calls were too few	0.001	0.067	0.072	0.000	0.546	0.000	2,048
	(0.032)	(0.250)	(0.259)				
Thought the calls were just right	0.003	0.133	0.138	0.000	0.687	0.000	2,048
	(0.056)	(0.340)	(0.345)				

Thought the calls were too many	0.000	0.031	0.023	0.000	0.175	0.000	2,048
	(0.000)	(0.175)	(0.150)				
Found calls useful	0.003	0.208	0.209	0.000	0.951	0.000	2,048
	(0.056)	(0.406)	(0.407)				
Attempted to authenticate E-tag maize	0.001	0.008	0.010	0.005	0.491	0.084	2,048
	(0.032)	(0.089)	(0.100)				
Successfully authenticated E-tag maize	0.000	0.007	0.008	0.002	0.708	0.038	2,043
	(0.000)	(0.083)	(0.090)				
Attempted to authenticate E-tag herbicide	0.001	0.043	0.036	0.000	0.290	0.000	2,048
	(0.032)	(0.204)	(0.186)				
Successfully authenticated E-tag herbicide	0.001	0.042	0.035	0.000	0.332	0.000	2,015
	(0.032)	(0.200)	(0.183)				

Notes: Standard deviations are reported in parentheses. P-values for differences in means are derived from two separate regressions. The first is a regression of the variable of interest on a dummy variable for being assigned to the encouragement group without including the discount group and comparing testing whether the coefficient on the encouragement dummy is significantly different from zero. The second is a regression of the outcome of interest on dummy variables for the encouragement treatment and the discount treatment, with tests for the equality of the encouragement and discount groups, as well as a test for whether the coefficient on the dummy variable for belonging to the discount group is significantly different from zero. All regressions include a dummy variable for not owning a phone and control for market location fixed effects to account for stratification and standard errors clustered at the village level. Means and regressions also weight observations according to their probability of selection into the sample based on whether the household owned a phone in the first CLE. Stars denote statistical significance as follows: * $p < 0.1$, ** $p < .05$, *** $p < 0.01$.

We will begin by describing the experiences of households with the interventions during the most recent season, the first season of 2017. We see that almost no households attended the community promotion meeting in the control group, and about 40 percent of households attended the meeting in the encouragement and discount groups, with slightly more households from the discount group attending. Approximately 40 percent of those who attended the meetings found the meetings helpful. That is not particularly high, although anecdotally, many households mentioned that the meetings were fun and entertaining. Almost no households in the control group reported receiving in interactive voice messages as well, and about 28 percent of the encouragement group and 31 percent of the discount group reported receiving the voice messages. There should not be differences in who received phone calls between the encouragement and discount groups, but it is possible that the discounts led people to have better recall. Further, 19 percent of the encouragement group and 20 percent of the discount group interacted with the voice messages. Once again, better recall may account for this difference. A few people (approximately 8 percent of households) also called the phone number back. When people called the number back, they were able to hear the same message and interact with the system in the same way. The largest proportion of people reported that the number of calls were just right; not too few, nor too many, and about a quarter of the encouragement and discount groups found the calls useful.

Looking at the discounts, almost nobody in the control group received the discount vouchers, 16 percent of people in the encouragement group (includes villages not assigned to the discount) received the vouchers, and about 30 percent of people received the maize and herbicide vouchers (this is not surprising as they were both given out together). Very few people actually redeemed the voucher, however. While 10 percent of the discount group redeemed the herbicide voucher, only 4 percent redeemed the hybrid maize voucher. This is consistent with the overall usage patterns we saw earlier; households are much more likely to use herbicide maize compared to hybrid maize. Unfortunately, as has been found with other similar authentication systems, almost nobody authenticated their E-tag herbicide or hybrid maize. Less than one percent of the control and encouragement groups attempted to authenticate E-tag maize, and only 2 percent of the discount group attempted. For E-tag herbicide, almost nobody in the control group attempted to authenticate in the control group, while 10 percent of people in the encouragement group and 15 percent of people in the discount group attempted to authenticate. In the discount group, enumerators often helped with authentication during the time when the discount vouchers were redeemed. We see similarly low rates of success in authenticating E-tag maize across groups, but we do see that 14 percent of the discount group was successfully able to authenticate their E-tag herbicide.

The encouragement interventions began in the second season of 2014. Here again we see that almost nobody in the control group attended the community meetings, but that about 42 percent of households in the encouragement and discount groups reported attending the meetings. Again approximately 40 percent of households found the meeting useful. The rate of receiving the phone calls is slightly lower in the first season at about 23 percent for the encouragement and discount groups, and this is likely because of technical issues still being worked out in the IVR system. Approximately 15 percent of households reported interacting with the system, and about 7 percent attempted to call the number back. Once again, many people thought that the number of calls was neither too few nor too many, and about 20 percent found the calls useful. Few people were attempting to authenticate and were successful at authenticating E-tag herbicide or maize. It seems that it took repeated meetings and calls (as well as the incentive from the discount) to get people to authenticate the E-tag products. This is a common problem that can severely limit the efficacy of such verification systems. If almost nobody is verifying, then companies cannot be held accountable for the products they are selling; the label has no meaning (apart from possibly a short-term signaling device).

We also have data directly from the IVR system. These data are not recall data, but they do differ in that we can only have information from the sample of households for which we had phone numbers. Table 7.4 displays summary statistics regarding the IVR messages that were sent in second season 2016 (the first round of messages). The rows present means of each of the outcome variables (one in each column), with standard deviations in parentheses below for variables that are not binary. Three different types of messages were sent in this season: an introduction to e-tag products before they reached markets, a message to say that E-tag products now had reached their markets, and a message to say that E-tag products and now E-tag maize seeds (which had taken longer to reach markets) were now available in their markets. The IVR messages introduced the verification scheme and allowed respondents to interact with the system by pressing 1 for a list of E-tag herbicide brands and by pressing 2 for a list of E-tag hybrid maize. More than one call was attempted for each message; one for the first, four for the second, and two for the third. The numbers presented here represent information for any of the attempts.

For the first message, 90 percent of calls went through (the number called was correct) to the phone numbers we had listed. For messages 2 and 3, all calls went through since only phone numbers for which calls went through in the first message were given messages 2 and 3. The proportion of calls that were finished (the respondent listened to the message until the end) was 27 percent for the first message, 36 percent for the second message, and 21 percent for the third message. The average call duration (the amount of time spent listening to the message) was almost two minutes for the first and second messages and was 0.7 minutes for the third message. Respondents had the option to call the number back if they had either missed it or wanted to listen to the message again. Approximately 3 percent of people called back for the first message, 4 percent for the second message, and 3 percent for the third message. Most people who called back listened to the full message. Importantly, the number of calls made was quite high; over 12,000 calls for the first message, 48,000 for the second message, and almost 26,000 for the third message. This indicates that the IVR messages were reaching a large number of people.

Table 7.4: Summary of VOTO Voice Messages in 2016

	(1) Introduction to EV products	(2) Notify farmers that EV products are in their market	(3) Notify farmers that EV products (including now seeds) are in their market
Proportion of calls that went through	0.903	1.000	1.000
Proportion of calls that were finished	0.272	0.361	0.213
Average call duration (minutes)	1.779	1.964	0.718
	(1.790)	(2.027)	(0.550)
Proportion of phone numbers who called the number back	0.034	0.038	0.030
Average duration of returned calls (minutes)	1.547	1.467	0.605
	(1.756)	(1.888)	(0.581)
Total number of calls sent to farmers	12,253	48,201	25,974
Total number of returned calls	420	1,855	786

In first season 2017, only one type of message was sent: a notification that E-tag and now also Ag-verify products were available in their markets. An SMS was also sent to phone numbers in the discount communities to notify farmers when the CPM teams would return to the LC1 to provide the coupon redemption money. Table 7.5 displays these statistics, with rows again presenting means of each of the variables, with standard deviations in parentheses below for variables that are not binary. Ninety-three percent of calls successfully went through, 36 percent of calls were completed to the end, and the average call duration was 1.7 minutes. To put this in context, across all of the types of messages and all languages in which messages were sent, the average duration of going through all of the options was 2.5 minutes. Farmers thus did listen to the bulk of each message. Only about 1 percent of people called back and if they did, they also appear to have listened to the full message. Forty-two percent of SMS messages successfully went through (this includes all encouragement only and discount group phone numbers).

Table 7.5: Summary of VOTO Voice Messages and SMS in 2017

	(1) Notify farmers that EV products/seeds are in their market	(2) SMS to notify farmers about discount vouchers
Proportion of calls that went through	0.926	
Proportion of calls that were finished	0.363	
Average call duration (minutes)	1.692 (1.977)	
Proportion of phone numbers who called the number back	0.011	
Average duration of returned calls (minutes)	1.685 (1.844)	
Proportion of SMS successfully sent		0.424
Total number of calls sent to farmers	33,115	
Total number of returned calls from farmers	1,182	
Total number of SMS sent to farmers		5,097

In general, the treatment was quite successful in reaching the study sample. Information was provided to a large number of people at the CPMs and through the IVR messages. This is particularly evidenced by the fact that more than half of the encouragement and the discount groups had heard of either E-tag or Ag-verify.

7.3 Impacts of the Encouragement Intervention on Take-up of Inputs and Farm Production

In this section, we will present estimates of the causal impacts of the encouragement intervention on outcomes of adoption of productivity enhancing agricultural inputs, as well as measures of farm productivity. These impacts are derived from estimating equation (1). In this section we report the impacts without household controls; impact estimates including household control variables are reported in Appendix A. The results are extremely robust to including these household level control variables.

The tables in this section follow a consistent format and can be interpreted in the same way. Each column represents a different outcome variable. The row labeled "Encouragement" shows the coefficient (impact estimate) of the individual being assigned to an encouragement community. It can be interpreted as the percentage point increase in the outcome variable resulting from being assigned to an encouragement community. The standard error of the impact estimate is reported below it and represents the statistical accuracy of the impact estimate (smaller numbers indicate higher accuracy). The asterisks represent whether the impact estimate is statistically significant; that is, whether the treatment group mean is statistically different from the control group mean. Three asterisks mean that the p-value is less than 0.01 so there is a less than one percent chance that means for the two groups are equal (1 percent level), two asterisks mean that the p-value is between 0.01 and 0.05 so there is a 1-5 percent chance that the two groups have equal means (5 percent level), and one asterisk mean that the p-value is between 0.05 and 0.1 so there is a 5 to 10 percent chance that means of the two groups are equal (10 percent level). The number of observations in the regression are reported in the row below the standard error, followed by the R^2 (which tells us how much of the variation in the outcome variable is explained by the variables included in the regression), and finally the average of the outcome variable in the control group is reported in the last row. This is reported so that the impact estimate can be put into context. For example, if the control group average of the outcome is 10 and the impact estimate is 5, that is a 5-percentage point increase in the outcome variable compared to the control group, which means the average in the treatment group is 15. This represents an increase in the outcome variable of 50 percent.

The e-verification intervention substantially increased the adoption of agricultural inputs. Table 7.4a reports impacts on input use in first season of 2017 and Table 7.4b reports impacts on input use from second season of 2016. We see that being assigned to an encouragement village increases the probability that a household has used hybrid maize by 3 percentage points in first season 2017, and by 1.8 percentage points in second season 2016. This is an increase of 27 percent in first season 2017, and of 24 percent in seconds season 2016, which are relatively large magnitudes. These effects are statistically significant at a less than 5 percent level. We see similarly sized impacts on households using any maize, which indicates that the increase in use of hybrid maize is coming from new farmers trying out hybrid maize. The percentage differences are lower for conventional maize, but this is unsurprising given that almost the entire sample is already growing maize. Being assigned to an encouragement village also increased the likelihood of using glyphosate herbicide and of using any herbicide by 4.3 and 3.7 percentage points in first season 2017, respectively, and these impacts are statistically significant at the one percent level. These are increases of 8-10 percent, which are also large given the larger base rates. In second season 2016 we do not see statistically significant increases in the use of herbicide. As mentioned earlier, we believe that households tend to purchase herbicide during the main growing season, the first season, and then tend to grow and invest less during second season. We also see that the encouragement causes a 1.8 percentage point increase in the use of inorganic fertilizer and a 3 percentage point increase of using any fertilizer again only in first season 2017 (a 20 percent increase). The same reasoning applies here: most input purchases are made during the first season. Fertilizer was not labeled, and so this effect appears to be a spillover effect. It could be the case that because people are investing in hybrid maize and herbicide, they also want the complementary benefits of fertilizer. Also possible is that when an individual is in a shop purchasing these other inputs, the shopkeeper advises them to also purchase fertilizer. A third possibility is that the intervention caused people to believe that the quality of all inputs available in their markets was higher. We will turn to this in detail in the next sub-section.

Table 7.4a: Reduced-form Impacts of the Encouragement Intervention on the Use of Inputs in First Season 2017

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Encouragement	0.030***	0.031***	0.043***	0.037***	0.018*	0.030**
	(0.010)	(0.011)	(0.013)	(0.013)	(0.010)	(0.012)
Number of observations	2,026	2,024	1,954	1,994	1,994	1,981
R-squared	0.163	0.118	0.425	0.430	0.163	0.243
Endline control group mean	0.105	0.888	0.438	0.454	0.102	0.150

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the average baseline outcome variables for first season 2014 and 2015, and for second season 2013 and 2015. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table 7.4b: Reduced-form Impacts of the Encouragement Intervention on the Use of Input in Second Season 2016

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Encouragement	0.018**	0.021*	0.003	0.007	0.001	0.000
	(0.009)	(0.012)	(0.013)	(0.013)	(0.008)	(0.010)
Number of observations	1,942	1,959	1,868	1,930	1,930	1,930
R-squared	0.169	0.146	0.404	0.408	0.131	0.157
Endline control group mean	0.076	0.808	0.350	0.372	0.062	0.100

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the average baseline outcome variables for first season 2014 and 2015, and for second season 2013 and 2015. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Next, we turn to the results of farm level outcomes, which are the outcomes we ultimately want to improve. In Table 7.5a and Table 7.5b we examine the number of crops grown by the household, the natural log[11] of the total yield from maize (reported in tons per acre), the natural log of the total value of farm production in millions of UGX per acre, and the net income from farm work (profits) in millions of UGX.[12] Column (1) to (4) are reported across all plots, and columns (5) to (7) are reported for the representative maize plot. The representative maize plot was a randomly selected plot out of all of the plots on which a household grew maize. This was only done for the most

[11] The variable is transformed using the natural log to lessen the impact of outliers in the data on the point estimate, which is an average of the treatment effect over all households.
[12] For the net income outcome, instead of using the natural log, which only transforms positive numbers, we use the inverse hyperbolic sine (HHS) transformation because it is also able to transform negative numbers. Net income can be negative.

recent season, which was first season 2017. These data were better collected; for example, GPS units were used to measure the area of the plot, whereas recall data were used for the other plots.

We do not see statistically significant impacts on number of crops grown, maize yield, value of farm production, or on net income (profits) in first season 2017. In second season 2016, the encouragement led farmers to grow more crops than the control group (statistically significant at the 5 percent level), although the magnitude is quite small.

Table 7.5a: Reduced-form Impacts of the Encouragement Intervention on Farm Production Outcomes in First Season 2017

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)	(5) Log maize yield in representative maize plots (tons per acre)	(6) Log value of farm production in representative maize plots (million UGX per acre)	(7) IHS* of net income of farm production in representative maize plots (million UGX)
Encouragement	0.038	-0.026	-0.010	0.161	-0.037	-0.011	-0.459
	(0.047)	(0.042)	(0.038)	(0.249)	(0.042)	(0.075)	(0.244)
Number of observations	2,039	1,641	1,932	1,962	1,611	1,638	1,616
R-squared	0.248	0.276	0.208	0.141	0.254	0.324	0.156
Endline control group mean	4.199	4.905	12.497	12.246	5.891	13.411	11.895

Notes: IHS* is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * $p < 0.1$, ** $p < .05$, *** $p < 0.01$.

Table 7.5b: Reduced-form Impacts of the Encouragement Intervention on Farm Production Outcomes in Second Season 2016

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)
Encouragement	0.109**	0.001	0.004	0.275
	(0.047)	(0.052)	(0.032)	(0.190)
Number of observations	2,034	1,498	1,864	1,888
R-squared	0.258	0.287	0.210	0.126
Endline control group mean	3.629	4.906	12.671	12.901

Notes: IHS* is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(4) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * $p < 0.1$, ** $p < .05$, *** $p < 0.01$.

7.4 Impacts of E-verification on Take-up of Inputs and Farm Production

In this section we present results of estimating equations (2) and (3); the instrumental variables regression model. In the first stage, we use the encouragement group dummy variable to predict whether the respondent has heard of E-tag/Ag-verify, and then in the second stage we use predicted knowledge of E-tag/Ag-verify products to estimate impacts on outcomes of herbicide, maize, and fertilizer use, as well as number of crops, yields, and net incomes. This will provide us with the main impact estimate of interest for this project: the impact of e-verified products on adoption and farm outcomes. Appendix A also contains the regressions that include household demographic controls.

The tables in this section follow a different format than those in the previous section, but within this section, the tables can all be read and interpreted in the same way. Each column represents a different final outcome variable (adoption and farm outcomes). The panel labeled "first stage" is the regression for Equation (2), where the outcome variable is whether an individual has heard of E-tag/Ag-verify. The row labeled "Encouragement" shows the coefficient (impact estimate) of the individual being assigned to an encouragement community on awareness of e-verification. It can be interpreted as the probability of being aware of E-tag/Ag-verify resulting from being assigned to an encouragement community. The standard error of the impact estimate is reported below it and represents the statistical accuracy of the impact estimate (smaller numbers indicate higher accuracy). The asterisks represent whether the impact estimate is statistically significant; that is, whether the treatment group mean is statistically different from the control group mean. Three asterisks mean that the p-value is less than 0.01 so there is a 1 percent chance or less that the two groups are equal (1 percent level), two asterisks mean that the p-value is between 0.01 and 0.05 so there is a 1 to 5 percent chance that means of the two groups are equal (5 percent level), and one asterisk mean that the p-value is between 0.05 and 0.1 so there is a 5 to 10 percent chance that the means of the two groups are equal (10 percent level). The row labeled "first stage F-statistic" reports a statistical test of whether Equation (2) is informative; that is, whether the encouragement assignment has a meaningful impact on awareness of E-tag/Ag-verify. If it did not, encouragement assignment would not be useful for estimating the eventual impacts of e-verification on the adoption and farm outcomes. F-statistics above 10 indicate that the encouragement intervention is meaningful and can be used to predict whether an individual has heard of E-tag/Ag-verify. It is from this estimation of Equation (2), we obtain predicted values of whether an individual has heard of E-tag/Ag-verify.

In the second panel, labeled "second stage" we report the output of Equation (3). Here, we use the predicted values of whether an individual has heard of E-tag/Ag-verify to estimate the impact of e-verification on adoption and farm outcomes. The row labeled "heard of e-tag – predicted" reports the impact estimate of the e-verification scheme on the outcome for that column. The coefficient can be interpreted as the percentage point increase in the outcome variable as a result of being aware of the E-tag/Ag-verify products. The standard error of the impact estimate is reported below it. Asterisks are interpreted in the same way.

The number of observations in the regression are reported in the next row, followed by the R^2 (which tells us how much of the variation in the outcome variable is explained by the variables included in the regression), and finally the average of the outcome variable in the control group is reported in the last row. This is reported so that the impact estimate can be put into context.

Table 7.6a and Table 7.6b display IV estimates of impacts on adoption of hybrid maize, any maize, glyphosate herbicide, any herbicide, inorganic fertilizer, and any fertilizer. First note that the first stage (the regression of the encouragement dummy variable on a dummy variable for whether the household has heard of E-tag/Ag-verify) is strong. Being assigned to an encouragement village increases the probability of knowledge of the E-tag/Ag-verify schemes by approximately 56 percentage points. This is a very large increase, and the coefficient is statistically significant at the one percent level. Additionally, the F-statistic for the first stage is large (over 1,200). This shows that the encouragement treatment is both a valid and a strong predictor of the uptake of EV.[13]

The results very much mirror those from the reduced form impacts of the encouragement intervention. We see that in first season 2017, similar to the reduced form impacts, adopting e-verified products leads to a 5.4 percentage point increase in the likelihood that a household uses hybrid maize and a 5.6 percentage point increase in the likelihood that a household grows any maize. Again, the results indicate that the difference in the use of hybrid maize is due to new households adopting this input. We also see that the likelihood of using glyphosate herbicide is increased by 7 percentage points, as is the likelihood of using any herbicide. Here, we also see impacts on the use of inorganic fertilizer and on the use of any fertilizer, with increases of 3 and 5 percentage points, respectively. The effect sizes from the IV specification are slightly larger than those of the reduced form. These estimates are local average treatment effects (LATE), representing the effect e-verification on outcomes for households that were induced to gain knowledge of e-verification as a result of being exposed to the encouragement intervention.

In second season of 2016, we see that the encouragement instrument is once again both strong and of the same magnitude and the F-statistic of approximately 1,100. The likelihood of knowing about E-tag/Ag-verified products is lower in second season 2016, at approximately 3-4 percentage points as a result of the encouragement, but this this is because it was the first season in which the schemes were introduced, so fewer people would have heard about them (this was also clear from section 7.1). In the second stage we see that households are more likely to use hybrid maize, and again, the difference is coming from the intervention inducing new farmers to adopt maize production. Once again we see that households are not statistically more likely to have used herbicide or fertilizer, and again, we believe this is because such purchases are made during the first growing season.

[13] The reason that the coefficients are slightly different across the different outcomes is due to the slightly different numbers of observations. For some households, endline values of these outcomes are missing.

Table 7.6a: Impact of E-verification on Use of Agricultural Inputs in First Season 2017, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
First stage						
Encouragement	0.567***	0.563***	0.560***	0.559***	0.562***	0.561***
	(0.016)	(0.016)	(0.017)	(0.016)	(0.016)	(0.016)
First stage F-statistic	1,224.1	1,197.2	1,115.7	1,150.5	1,178.2	1,170.6
Second stage						
Heard of E-tag - predicted	0.054***	0.056***	0.077***	0.066***	0.032*	0.053***
	(0.017)	(0.018)	(0.023)	(0.023)	(0.017)	(0.020)
Number of observations	2,024	2,023	1,954	1,993	1,993	1,981
R-squared	0.042	0.017	0.136	0.143	0.061	0.057
Endline control group mean	0.105	0.888	0.438	0.454	0.102	0.150

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the average baseline outcome variables for first season 2014 and 2015, and for second season 2013 and 2015. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table 7.6b: Impact of E-verification on Use of Agricultural Inputs in Second Season 2016, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
First stage						
Encouragement	0.563***	0.562***	0.563***	0.559***	0.561***	0.561***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
First stage F-statistic	1,158.9	1,145.6	1,082.9	1,111.1	1,135.0	1,129.8
Second stage						
Heard of E-tag - predicted	0.031**	0.037*	0.005	0.012	0.002	0.001
	(0.015)	(0.020)	(0.023)	(0.023)	(0.014)	(0.017)
Number of observations	1,942	1,959	1,868	1,930	1,930	1,930
R-squared	0.030	0.007	0.111	0.108	0.049	0.030
Endline control group mean	0.076	0.808	0.350	0.372	0.062	0.100

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the average baseline outcome variables for first season 2014 and 2015, and for second season 2013 and 2015. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Next, we look at outcomes relating to farm production and income. Tables 7.7a and 7.7b report the IV impact estimates for the number of crops grown, log maize yield, log total value of farm production, and net income from farm work.

For first season 2017 and for second season 2016, the first stage is once again both strong and relevant, with again a very similar magnitude to the previous tables. Once again, the impact coefficients are slightly larger than the reduced form coefficients, but the message is the same. There is a marginally statistically significant negative impact on net income on the representative maize plot in first season 2017. It is possible that this was due to adoption of the inputs; they all involved costs and the costs may have outweighed the income from using them. Hybrid maize was a new product for many, herbicide does not improve yields, only reduces labor time, and fertilizer was not tagged so may not have been of higher quality. There is evidence that more crops were grown in second season 2016 but yield and the value of production were not statistically affected by e-verification. Net income is also not statistically significant, but we note that the point estimate is large.

Table 7.7a: Impact of E-verification on Farm Production Outcomes in First Season 2017, IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)	(5) Log maize yield in representative maize plots (tons per acre)	(6) Log value of farm production in representative maize plots (million UGX per acre)	(7) IHS* of net income of farm production in representative maize plots (million UGX)
First stage							
Encouragement	0.565***	0.589***	0.569***	0.563***	0.592***	0.596***	0.593***
	(0.016)	(0.018)	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)
First stage F-statistic	1,191.0	1,055.2	1,187.8	1,165.5	1,052.0	1,116.4	1,099.5
Second stage							
Heard of E-tag - predicted	0.064	-0.045	-0.017	0.286	-0.062	-0.019	-0.775*
	(0.080)	(0.069)	(0.064)	(0.426)	(0.068)	(0.120)	(0.397)
Number of observations	2,035	1,641	1,932	1,962	1,611	1,638	1,616
R-squared	0.086	0.025	0.008	0.003	0.004	0.002	-0.000
Endline control group mean	4.199	4.905	12.497	12.246	5.891	13.411	11.895

Notes: IHS* is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table 7.7b: Impact of E-verification on Farm Production Outcomes in Second Season 2016, IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income from farm work (million UGX)
First stage				
Encouragement	0.563***	0.556***	0.563***	0.561***
	(0.016)	(0.018)	(0.017)	(0.017)
First stage F-statistic	1,192.3	921.5	1,065.3	1,087.7
Second stage				
Heard of E-tag - predicted	0.185**	0.003	0.007	0.491
	(0.081)	(0.089)	(0.054)	(0.326)
Number of observations	2,030	1,498	1,864	1,888
R-squared	0.076	0.036	0.010	0.008
Endline control group mean	3.629	4.906	12.671	12.901

Notes: IHS* is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(4) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Overall, we find that the e-verification initiative causes a substantial increase in the adoption of productivity-enhancing inputs, particularly of hybrid maize and glyphosate herbicide, which were the two products that were tagged. However, e-verification did not lead to increases in farm incomes or profits in general. One reason could be that the program may have had no effects on farm incomes because glyphosate herbicide does not improve yields, but rather, saves labor time. These labor-saving benefits could have led to increases in farm profits, but this was not a result supported by these data. It is difficult to collect data on farm inputs and outputs; some data is missing and needs to be imputed, recall data is notoriously difficult (it is difficult for farmers to remember plot sizes, harvest sizes with precision), and aggregating these sources of noise in the data into yields and net income then becomes problematic.

7.5 Impacts of E-verification on Beliefs Regarding Quality of Inputs

The tables in this section follow the same format and can be interpreted in the same way as the tables in the previous section and estimate IV specifications (Equations 2 and 3). We now turn to the impacts of the e-verification initiative on beliefs regarding input quality. Our hypothesis is that one of the reasons that many farmers do not take up inputs is because they believe the inputs are of low quality, which can reduce expected returns enough that they no longer justify the cost. Another way to view this is that farmers may not take up inputs because they believe that a substantial share of inputs on the market are counterfeit or adulterated and they cannot afford the risk of low or even negative returns if the input they purchase turns out to be fake. If it is indeed the case that farmers were not purchasing inputs because of perceptions of low quality, then the e-verification initiative may have shifted beliefs regarding the veracity of these inputs. This avoidance behavior is document in the baseline report (before the intervention), where 80 percent of farmers report that they have avoided purchasing maize seed or, separately, herbicide, in the past because they believed the inputs were counterfeit/adulterated (Ashour et al, 2015).

The e-tag and Ag-verify labels aimed to improve farmers' confidence in the fidelity of the inputs they purchased by acting as a form of product assurance. The SMS received after submitting the code on the scratch labels provided information regarding the brand, package size and expiration date. This information, and the presence of the e-tag or Ag-verify label itself, should reassure farmers that the product they purchased corresponds to the description on the package. The effect of this product assurance strategy on farmers' beliefs about the quality of inputs in the market depends on several factors. Farmers may believe that the scheme is increasing overall input quality in the market by reducing the likelihood of counterfeiting between the moment of packaging and sale. However, if farmers believe that most counterfeiting occurs before the input is packaged, or that an e-verified brand is a low-quality brand due to its formulation, then e-verification may not improve farmers' perceptions about the quality of the e-verified input, or inputs in general. In addition, the presence of the e-tag and Ag-verify system and the encouragement information campaigns run for this study may increase awareness about input counterfeiting and adulteration as a problem and may temporarily contribute to beliefs that inputs are low quality.

We again focus on the IV regressions in order to estimate the impact of the e-verification intervention (rather than our messaging) on beliefs regarding quality of inputs. We elicit farmer beliefs regarding quality of inputs in two parts of the survey. During the household survey, we ask farmers whether they believe conventional maize, hybrid maize, fertilizer, and herbicide are ever counterfeit/adulterated. These questions are only asked of farmers who had purchased the input in the past two growing seasons. As a result, the beliefs they are expressing reflect beliefs over both season one 2017 or season two 2016, so the responses reflect the beliefs of farmers with recent experience purchasing the input. These farmers are likely a selected group in that they have different demographic characteristics compared to the average farmer (Ashour et al, 2018), which could influence the way they answer these questions. Nonetheless, the perspective of farmers with recent experience is useful in part because they may be better informed than their peers. In Table 7.8 below, we report the impact estimates of e-verification on the likelihood that farmers believe that the quality of these inputs is ever lowered due to counterfeiting or adulteration. We see that e-verification led farmers to believe that it was more likely that maize (conventional or hybrid) quality was ever lowered by counterfeiting/adulteration, by 10 and 16 percentage points, respectively, and these coefficients are statistically significant. These differences in beliefs represent an 18 and 22 percent change, respectively, which are quite substantial effect sizes. We find no effect of e-verification on beliefs about whether inorganic fertilizer or glyphosate herbicide are ever counterfeited/adulterated; point estimates on inorganic fertilizer and glyphosate herbicide are negative but are not statistically significant.

Table 7.8: Impacts on Qualitative Beliefs about Input Quality in 2016-2017, IV Specifications

	(1) The quality of conventional maize is purposely lowered by counterfeiting/ adulteration	(2) The quality of hybrid maize is purposely lowered by counterfeiting/ adulteration	(3) The quality of glyphosate herbicide is purposely lowered by counterfeiting/ adulteration	(4) The quality of inorganic fertilizer is purposely lowered by counterfeiting/ adulteration
First stage				
Encouragement	0.549***	0.646***	0.591***	0.638***
	(0.030)	(0.068)	(0.021)	(0.060)
First stage F-statistic	338.1	89.5	826.6	114.3
Second stage				
Heard of E-tag - predicted	0.095***	0.161**	-0.006	-0.008
	(0.036)	(0.081)	(0.025)	(0.098)
Number of observations	718	235	978	246
R-squared	0.024	0.111	0.008	0.026
Endline control group mean	0.518	0.727	0.857	0.534

Notes: The outcome variables are specifically for the sample of HHs that purchased the inputs in seasons 2016 and 2017. Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for baseline outcome variables in 2013-2014. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

These results suggest that awareness of e-tag products causes farmers to update their prior beliefs in such a way that they believe it is more likely than before that conventional or hybrid maize seed may be counterfeit or adulterated.

We are able to further understand patterns of farmer beliefs by separating beliefs regarding Kakasa and non-Kakasa inputs. We ask farmers, for each input (hybrid maize, glyphosate herbicide, and inorganic fertilizer), whether none, less than half, half, more than half, or all is counterfeit/adulterated, and we do so separately for Kakasa and non-Kakasa products. Table 7.9

reveals an interesting pattern and suggests that e-verification shifted farmer beliefs regarding input quality for Kakasa and non-Kakasa products in different ways. The outcome variable is a dummy variable for whether the particular input is considered to have half or more products counterfeit/adulterated. For Kakasa hybrid maize and herbicide, farmers report that it is less likely that half or more of the input is counterfeit/adulterated by 13 and 5 percentage points, respectively. These differences represent a change in beliefs of more than 100 percent reduction in the likelihood that Kakasa maize is counterfeit, and a reduction of almost 100 percent in the likelihood of Kakasa herbicide is counterfeit. Since hybrid maize is purchased by fewer people, the base rate of the belief that the likelihood of the maize being of low quality may be due to less experience and the fact that maize may be easier to counterfeit (it is often sold in open bags) and farmers do not often even know if they use conventional or hybrid maize. Farmers have more experience with herbicide and the bottles are sold with a tamper-proof plastic seal, which may explain why the base rate for the belief that herbicide quality is low is much lower. For non-Kakasa products, farmers report that it is more likely that half or more of the input is counterfeit/adulterated, and the changes represent a change in beliefs that more than half of the product is of low quality. Thus, while e-verification may not have shifted beliefs substantially for glyphosate herbicide overall, farmers form separate beliefs regarding the quality of products that are tagged and those that are not tagged, and even those beliefs shifted in opposite directions as a result of the e-verification scheme.

For fertilizer, e-verification increased the amount of non-Kakasa fertilizer that is believed to be counterfeit/adulterated, though the effect is not significant. The estimated effect on beliefs about the share of Kakasa fertilizer that is counterfeit/adulterated is positive but is not statistically significant. Fertilizer was not part of the e-verification program and thus it is not surprising that there are no effects on Kakasa fertilizer. The decrease in the amount of Kakasa products believed to be counterfeit for both tagged inputs suggests that the e-verification intervention did indeed lead farmers to be more confident that the inputs were genuine. For non-Kakasa inputs, it appears that the e-verification intervention led to more awareness among farmers regarding issues on counterfeiting/adulteration or that the intervention may have led to farmers believing that products that were not tagged were of lower quality. In sum, these results point towards product assurance programs being able to shift farmers' beliefs regarding input quality.

Table 7.9: Impacts on Beliefs about Input Quality in 2016-2017, IV Specifications

	(1) More than half of non-Kakasa hybrid maize is adulterated or counterfeit	(2) More than half of Kakasa hybrid maize is adulterated or counterfeit	(3) More than half of non-Kakasa glyphosate herbicide is adulterated or counterfeit	(4) More than half of Kakasa glyphosate herbicide is adulterated or counterfeit	(5) More than half of non-Kakasa inorganic fertilizer is adulterated or counterfeit	(6) More than half of non-Kakasa conventional maize is adulterated or counterfeit
First stage						
Encouragement	0.711***	0.703***	0.613***	0.613***	0.612***	0.674***
	(0.079)	(0.072)	(0.023)	(0.023)	(0.040)	(0.101)
First stage F-statistic	81.0	95.6	717.6	730.0	230.0	44.5
Second stage						
Heard of E-tag - predicted	0.065	-0.131**	0.078**	-0.049***	0.105*	0.133
	(0.107)	(0.050)	(0.032)	(0.013)	(0.057)	(0.101)
Number of observations	177	177	840	840	389	130
R-squared	0.021	0.023	0.007	0.017	0.005	0.129
Endline control group mean	0.250	0.084	0.276	0.056	0.299	0.173

Notes: The outcome variables are specifically for the sample of HHs that purchased the inputs in seasons 2016 and 2017. Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for baseline outcome variables in 2013-2014. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * $p < 0.1$, ** $p < .05$, *** $p < 0.01$.

The second method by which we measured beliefs regarding the quality of inputs was through the quantitative ‘games’ that were undertaken in groups in the evening after the household surveys were completed. We asked farmers about hybrid maize, glyphosate herbicide, and inorganic fertilizer. The question asked was “Imagine ten farmers like you, went to [FARMER’S MARKET LOCATION] and purchased bottles of [INPUT] from the local agricultural inputs shop. Out of those ten bottles/bags of [INPUT] that the ten farmers purchased, how many do you expect are not genuine (that is, either the quality is lowered by mixing with fake or inferior product, or, it is completely replaced by fake product)? What number do you think is most likely?” Thus, the questions combined counterfeiting and adulteration and as such, can be interpreted as farmer beliefs regarding both aspects of quality.

Table 7.10 displays the results of estimating impacts on beliefs using equations (2) and (3) for these questions. We do not see an impact of the e-verification initiative on farmer beliefs regarding the quality of inputs when measured using this method. The effect sizes are quite small, and they are not measured very precisely – the standard errors are quite large.

Comparing tables 7.8 and 7.10, a noticeable pattern is that although some coefficient estimates are not statistically significant, for the various inputs they follow the same direction. For herbicide, farmers believe that fewer bottles of Kakasa herbicide are counterfeit or adulterated. The pattern for fertilizer is similar to that of herbicide. For maize, farmers believe that more bags are counterfeit/adulterated and they are more likely to believe that maize (conventional or hybrid) is counterfeit/adulterated. These patterns suggest that the e-verification intervention contributed to some overall changes in beliefs regarding the quality of inputs.

Table 7.10: Impacts on Quantitative Beliefs about Input Quality in 2016-2017, IV Specifications

	(1) Number of bags of hybrid maize (out of 10) counterfeit/adulterated	(2) Number of herbicide bottles (out of 10) counterfeit/adulterated	(3) Number of bags of fertilizer (out of 10) counterfeit/adulterated
First stage			
Encouragement	0.559***	0.559***	0.559***
	(0.016)	(0.016)	(0.016)
First stage F-statistic	1,192.0	1,173.3	1,166.7
Second stage			
Heard of E-tag – predicted	0.064	-0.027	-0.098
	(0.105)	(0.098)	(0.121)
Number of observations	2,060	2,060	2,060
R-squared	0.011	0.012	0.005
Endline control group mean	4.147	3.989	2.975

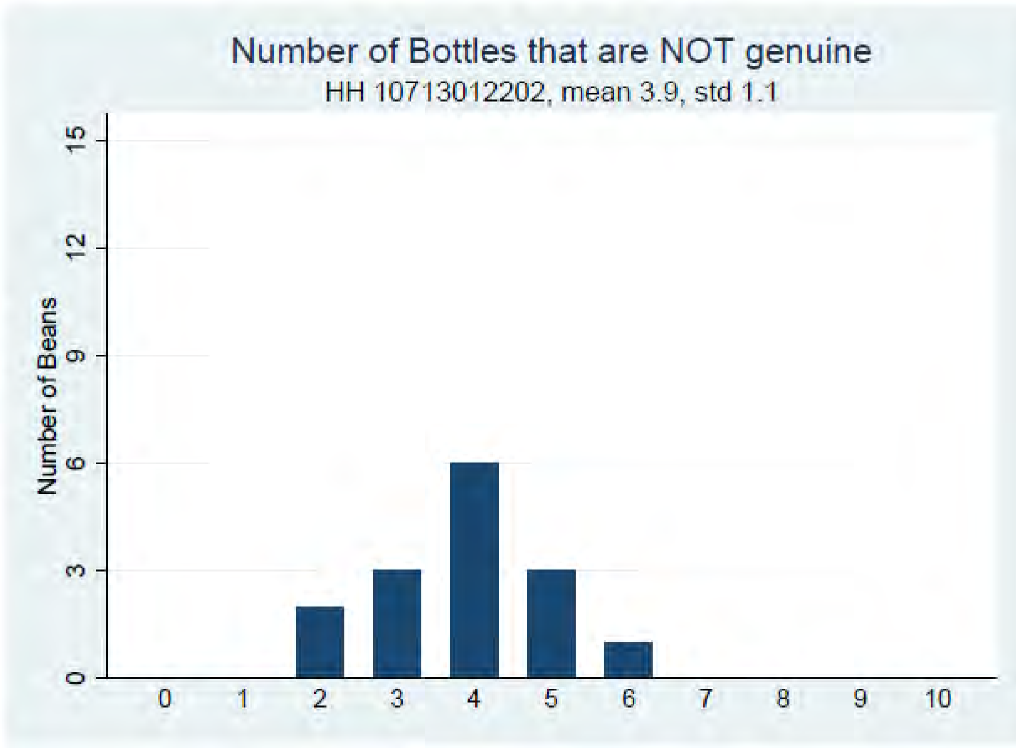
Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for baseline outcome variables in 2013-2014. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Once we had asked respondents about the number of bottles/bags (out of 10) of the input they believed was most likely to be counterfeit/adulterated, we conducted a structured elicitation game with farmers to ask them to show us the distribution of their beliefs. In the game, we showed farmers a large card with spaces numbered from 0-10 (see Figure 7.1 below). The spaces represented the chances that zero bottles/bags were counterfeit, one bottle/bag, two bottles/bags, etc. We also gave respondents 15 buttons, and asked respondents to place the buttons in the spaces according to how many bottles/bags they considered to be counterfeit. The more likely a number was, the more buttons should be placed in that space. The result could have looked like that in Figure 7.2 (this is an example). If the respondent believed that 4 out of 10 bottles were most likely to be counterfeit/adulterated, the most buttons would be placed in that space. Generally, we would expect that spaces 3 and 5 would have fewer buttons, spaces 2 and 6 even fewer, etc. This was a difficult exercise for respondents, and enumerators reported that many respondents were confused as the concept of probabilities and distributions was not something with which they had much experience. Some respondents are reported as having given up, or as having placed buttons randomly and sometimes in ways that did not make sense (for example, having a bi-modal distribution or their answer to ‘most likely’ not being the same number as the space with the most buttons). Although these data are noisy, we report our findings to be transparent, and we have not removed observations whose distributions appear different than we expected in order to not bias the results.

Figure 7.1: Beliefs Distribution Elicitation Tool

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Figure 7.2: Example Beliefs Distribution



Below, we show the distributions of the beliefs of respondents on average (the average number of buttons placed in each space over the ten spaces) for each of the three main inputs (hybrid maize, glyphosate herbicide, and inorganic fertilizer) for the treatment versus the control group.

Figure 7.3 displays the distributions for hybrid maize, Figure 7.4 for herbicide, and Figure 7.5 for fertilizer, respectively. If the encouragement promotion reduced the share of input purchases that respondents believe are counterfeit (through a reduced form effect inducing more take-up of e-tag products), we would expect to see that respondents in the encouragement group would place more buttons in lower numbered spaces compared to the control group. For hybrid maize, we do not see a particular shift in general: more respondents in the control group place buttons in the space for zero bags being counterfeit/adulterated but they also place more buttons in the space for five. Encouragement respondents place more buttons in some of the other spaces, but they are on both sides of the distribution. In general, there does not seem to be too much difference between the two distributions. We conduct a Mann-Whitney test for equality of the two distributions and cannot reject the hypothesis that the two distributions are the same; the p-value is 0.65. Measured with this method, we cannot say that there is evidence that the encouragement promotion changed the distribution of beliefs differentially between the encouragement and control groups.

For herbicide (Figure 7.4) the distribution for the encouragement group looks a bit as if the respondents placed more buttons in lower spaces. It is not immediately clear, but the Mann-

Whitney test reports a p-value 0.12, which, while not statistically significant at conventional levels, is relatively low. We should not conclude definitively that the distribution of beliefs shifted, but there is suggestive evidence that respondents in the encouragement group may have believed it was more likely that that less herbicide was counterfeit/adulterated.

For fertilizer (Figure 7.5) the distribution of beliefs does look different between the treatment and control groups. Respondents in the encouragement group placed more buttons on average in the spaces with lower numbers. It appears that beliefs have shifted for fertilizer. The Mann-Whitney test reveals that this difference is statistically significant (p-value 0.01). Because fertilizer was not tagged, this result may seem puzzling. Once again, we cannot prove any hypotheses definitively, but it is possible that, given that the encouragement intervention did induce further take up of fertilizer, the encouragement group was more likely to take up fertilizer *because* they believed it to be less likely to be counterfeit. This is consistent with the interpretation on the results presented earlier; the encouragement messaging may have improved beliefs in general regarding the quality of inputs and thus improved take up of inputs that were and were not tagged. The results do contradict each other somewhat, but because of the different methods of elicitation of beliefs and the different outcomes, it is difficult to say exactly what the mechanisms were. Table 7.11 reports the p-values of the Mann-Whitney tests for equality of distributions for each of the three inputs. Recall that p-values below 0.01 indicate a less than one percent probability that the means are equal, p-values between 0.01 and 0.05 indicate a 1 to 5 percent probability that the means are equal, and p-values between 0.05 and 0.1 indicate a 5 to 10 percent probability that the means are equal.

Figure 7.3: Distribution of Beliefs regarding quality of Hybrid Maize

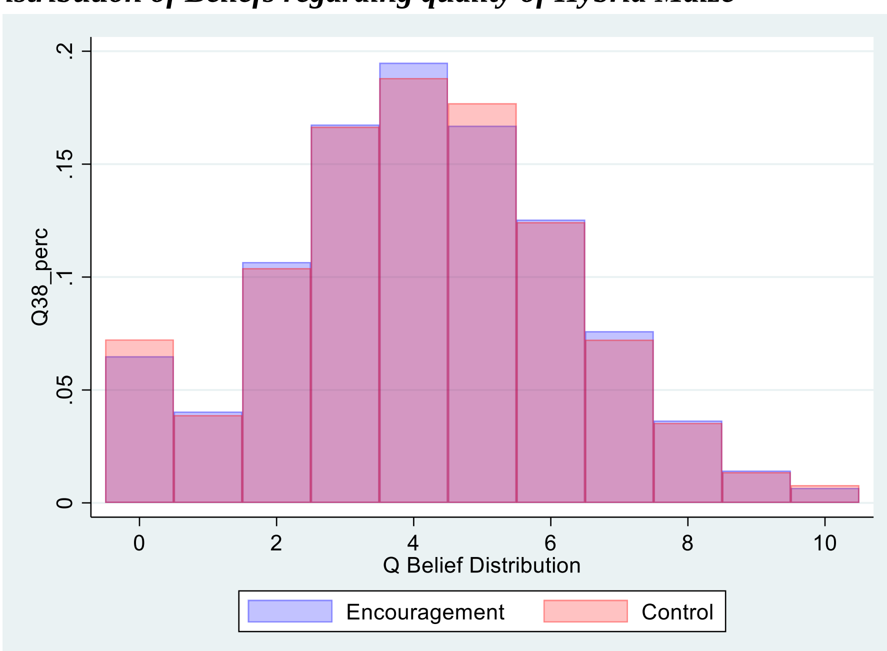


Figure 7.4: Distribution of Beliefs regarding quality of Herbicide

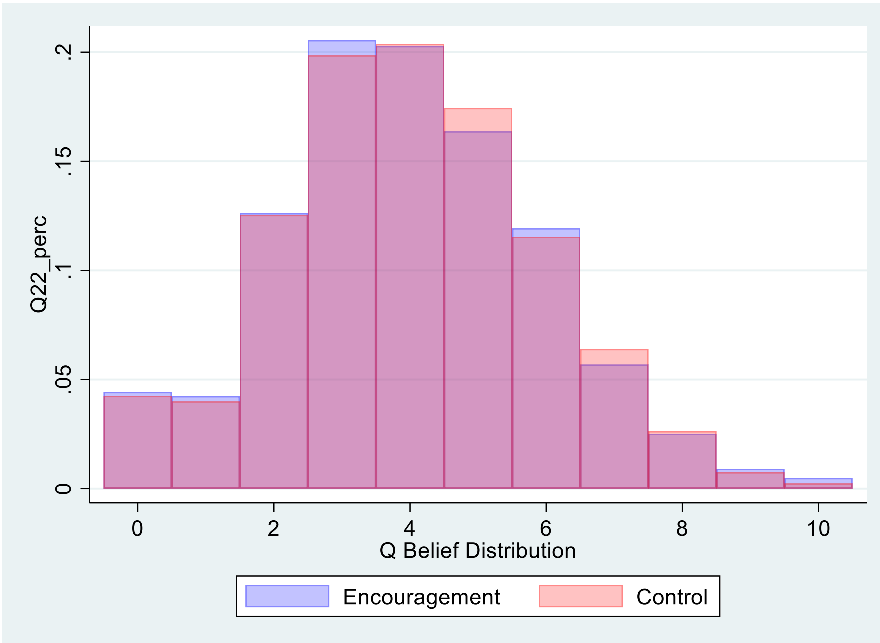


Figure 7.5: Distribution of Beliefs regarding quality of Fertilizer

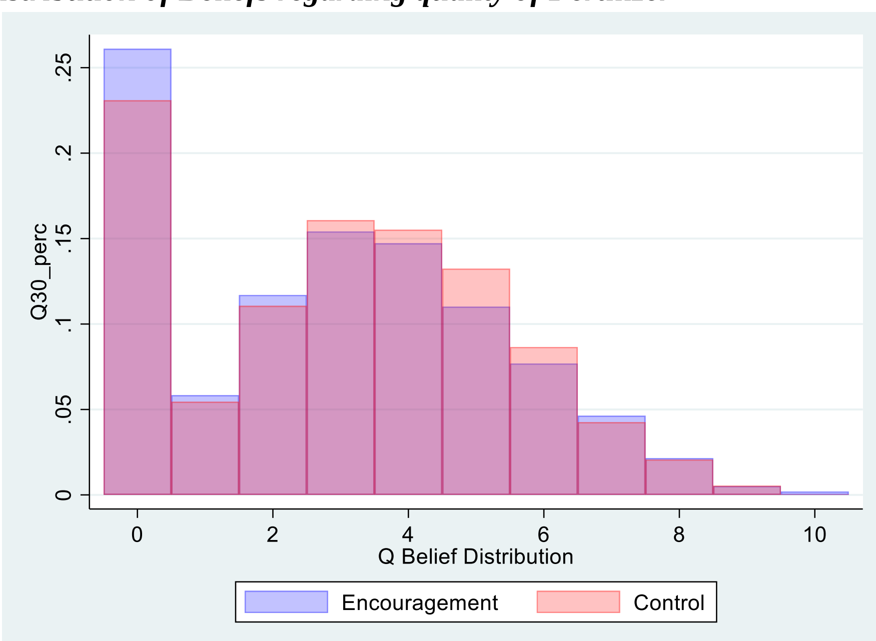


Table 7.11: P-values from Mann-Whitney Tests for Equality of Distributions

	Hybrid Maize	Herbicide	Fertilizer
P-value	0.65	0.12	0.01

Taken together, the results in this section show that the encouragement messaging did shift beliefs regarding the quality of agricultural inputs in local markets. In particular, farmers in the encouragement group believed that Kakasa inputs (hybrid maize and glyphosate herbicide) were less likely to be of low quality and that non-Kakasa inputs were more likely to be of low quality. The qualitative questions show more clearly how beliefs changed. There is also a shift in beliefs for fertilizer, which was not tagged. We hypothesize that the Kakasa campaign may have drawn attention to the problem of counterfeit/adulterated inputs, and thus non-Kakasa products are seen as likely to be of lower quality, but it may also have given rise to the notion that the quality of inputs in general may be improving. This may be one reason we see that respondents believe that less fertilizer is likely to be counterfeit/adulterated and why we see impacts of the intervention on take up of fertilizer as well.

7.6 Heterogeneous Impacts of E-verification on Input Adoption and Farm Outcomes

In this section we will explore whether the impacts of the e-verification intervention differed for different types of households or farmers. For individual farmers we focus on the characteristics of the primary agricultural decision maker (PADM). We separately identify impacts for male and female PADMs, PADMs who have completed primary school or more and PADMs who have not completed primary school, households considered to be poor (belonging to the lowest quintile of an asset index) and not poor, PADMs who are more or less risk-averse, PADMs who are ambiguity averse and not ambiguity averse, and for PADMs who have a phone or do not have a phone. For each of these characteristics, we examine heterogeneity of impacts on take up of inputs as well as farming outcomes.

Female PADMs may have been impacted by the program differently because they are less likely to purchase inputs, grow different crops than male PADMs or have less control over other resources needed to take advantage of high-quality inputs. More educated PADMs may be able to adjust their behaviours more easily than those who are less educated, as they may be able to better internalize information or interact with the e-verification system and thus may purchase more inputs or have better ability to use the inputs to improve farm outcomes. Poorer households may not be able to afford purchasing inputs, so the e-verification platform may not be of as much use to them. More risk-averse farmers may be less likely to try new products, including Kakasa products. More ambiguity-averse farmers may either be less likely to try new products if they are not sure about how likely the new products are to improve outcomes, or they may be more likely to try new products if they believe that tagged products reduce ambiguity. Finally, PADMs who own a phone may be more able to access the e-verification platform since a phone is required to verify an input and may have had greater exposure to the SMS component of the encouragement platform.

To estimate differential impacts by PADM characteristics, we run separate instrumental variable regressions (e.g., one for female PADMs and one for male PADMs) – Equations (2) and (3). We then test whether the coefficient on the predicted value of the e-verification intervention from the second stage is statistically different between the two groups. In the tables below, we report the treatment effect for the full sample reported earlier in the top panel, and the treatment effect for both subgroups as well as a p-value from the test of whether the coefficients for the two groups are statistically equal in the bottom panel. If this p-value is less than 0.05, we conclude that there is a statistically significant difference in impacts between the two groups.[14] Note that, when comparing the impact on the overall sample and the two sub-samples being compared, the coefficients will not always add up to the coefficient estimate for the full sample, nor add up to a weighted average of the two groups based on sample sizes. This is because the comparison group for the full sample is different from the comparisons for the two sub-samples. The comparison for the full sample is between all households who received the encouragement interventions and all households who did not get the encouragement interventions. The comparison for the two sub-samples is between the sub-sample who received the encouragement interventions and the sub-sample who did not (for example, we compare female PADMs who did receive the encouragement to female PADMs who did not receive the encouragement, and we compare male PADMs who did receive the encouragement with PADMs who did not receive the encouragement). We need to run these regressions separately instead of interacting the treatment variable (encouragement) with the characteristic to see if the impacts are different. We need to do this because we would have two endogenous variables (being aware of e-verified products and the characteristic, for example, gender) and only one exogenous variable – being assigned to an encouragement village. One must have at least as many exogenous variables as endogenous variables, which we do not have. It would not be correct to use the encouragement assignment and the encouragement interacted with the characteristic in the first stage, since the characteristic (and thus the interaction) is not exogenous.

In this section, the tables can be read and interpreted as follows. Each column again represents a different ultimate outcome (adoption, farm outcomes). The first panel shows the same IV impact estimates from previous sections (the coefficients and standard errors), but only the second stage (Equation 3), as this is the impact estimate to which we are interested in comparing the two sub-samples to and we have already reported the first stage above. In the next panel, we report the impact estimates (coefficients and standard errors) for the two sub-samples for which we are interested in differences. These are also impact estimates from Equation (3) and we do not report the coefficient from Equation (2). These coefficients can be interpreted as the percentage point

[14] The p-values for the test of the difference in treatment effects between the two subgroups are based on bootstrapped standard errors with 1,000 replications.

increase in the outcome variable as a result of being aware of E-tag/Ag-verify *only for that particular sub-group*. The p-value of the difference in coefficients is the statistical difference between the coefficients for the two different groups. Recall that the asterisks represent whether the impact estimate is statistically significant; that is, whether the treatment group mean is statistically different from the control group mean. Three asterisks or a p-value less than 0.01 mean there is a 1 percent chance or less that the two groups are statistically different (1 percent level), two asterisks or a p-value between 0.01 and 0.05 mean there is a 1-5 percent chance that the two groups are statistically different (5 percent level), and three asterisks or a p-value between 0.05 and 0.1 mean there is a 10 percent chance that the two groups are statistically different (10 percent level). In the third panel we report average values of the outcome variables for four different groups: the control and treatment groups of the first sub-group (for example, female PADMs), and the control and treatment groups of the second sub-group (for example, male PADMs).

We begin with heterogeneous impacts by gender. Table 7.11a reports impacts of the e-verification intervention on take-up of inputs for female and male PADMs for first season 2017 and Table 7.11b reports impacts for second season 2016. One finding that the tables reveal is that input use is higher among male PADMs compared to female PADMs. This result may not only capture gender but may also capture other aspects such as income and education level, as households with female PADMs may be poorer and have lower education levels. These results mirror prior results that overall, impacts are larger in first season 2017 than in second season 2016, where we find more take up as a result of e-verification for the full sample. Overall, there are positive impacts on use of some inputs for both female and male PADMs relative to the control group in each season, but there are some differences in the patterns of these effects. In first season 2017, e-verification has positive impacts for male PADMs on take up of hybrid maize seed, inorganic fertilizer and any fertilizer, while we do not find any positive impact on take up of these inputs for female PADMs. Interpretation of this pattern of results requires some care because we cannot reject that e-verification had equal impacts on take-up of these inputs for male and female PADMs (see P-value of difference in coefficients), though the difference in impact on take up of inorganic fertilizer is weakly significant. Nonetheless, this pattern of results suggests that the positive impact on take-up of these inputs seen for the sample overall are likely driven by the adoption behavior of male PADMs, particularly where the estimated impact on female PADMs is small (as for hybrid maize and inorganic fertilizer). It is important to keep in mind, though, that these tests have weaker power to identify impacts on female PADMs because there are substantially fewer female PADMs than male PADMs. Also, we find large positive impacts of e-verification on take up of any maize, glyphosate herbicide and any herbicide by female PADMs relative to female PADMs in the control group. In second season 2016, e-verification has a significant impact on take-up of hybrid maize seed by male PADMs and any maize seed by female PADMs. There are no other significant impacts and no statistically significant differences in impacts between the two groups. One reason for this may be that there are substantially fewer female PADMs compared to male PADMs. We cannot conclude definitively then, that male and female PADMs were not differentially impacted by the e-verification intervention.

Table 7.11a: Heterogeneous Impacts of E-verification on the Use of Agricultural Inputs by PADM Gender in First Season 2017, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Full Sample	0.054***	0.056***	0.077***	0.066***	0.032*	0.053***
	(0.017)	(0.018)	(0.023)	(0.023)	(0.017)	(0.020)
Female PADM	0.017	0.084**	0.105**	0.092**	-0.026	0.049
	(0.027)	(0.033)	(0.048)	(0.044)	(0.035)	(0.041)
Male PADM	0.083***	0.035	0.087***	0.093***	0.077***	0.083***
	(0.025)	(0.023)	(0.029)	(0.029)	(0.023)	(0.027)
P-value of difference in coefficients	0.181	0.360	0.813	0.996	0.076	0.599
Number of observations - Female PADM	685	692	665	674	674	674
Number of observations - Male PADM	1,264	1,266	1,229	1,248	1,248	1,246
Endline control group mean - Female PADM	0.070	0.859	0.365	0.386	0.085	0.128
Endline control group mean - Male PADM	0.125	0.907	0.478	0.492	0.110	0.163

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table 7.11b: Heterogeneous Impacts of E-verification on the Use of Agricultural Inputs by PADM Gender in Second Season 2016, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Full Sample	0.031**	0.037*	0.005	0.012	0.002	0.001
	(0.015)	(0.020)	(0.023)	(0.023)	(0.014)	(0.017)
Female PADM	-0.010	0.109**	0.062	0.060	-0.024	-0.033
	(0.023)	(0.044)	(0.050)	(0.045)	(0.027)	(0.034)
Male PADM	0.049**	0.007	0.009	0.013	0.026	0.029
	(0.022)	(0.026)	(0.031)	(0.031)	(0.020)	(0.024)
P-value of difference in coefficients	0.211	0.174	0.493	0.544	0.267	0.269
Number of observations - Female PADM	658	666	634	652	652	652
Number of observations - Male PADM	1,221	1,229	1,174	1,218	1,218	1,218
Endline control group mean - Female PADM	0.041	0.799	0.277	0.298	0.055	0.092
Endline control group mean - Male PADM	0.097	0.814	0.387	0.410	0.067	0.106

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Next, we examine differential impacts on male and female PADMS with regards to farm outcomes such as yields, value of farm production, and net income. Table 7.12a displays impact estimates for male and female PADMs as well as for the full sample for first season 2017, and Table 7.12b for second season 2016. We first observe that in the control groups, the outcomes do not differ much by gender of the PADM. As with the full sample results, there are not many statistically significant effects, and this is likely due to the noise with which these outcomes are generally measured. For first season 2017, we see that the negative impact on net income is driven by female PADMs (on the order of approximately 2 million UGX). It is not clear what might be driving this effect. In second season 2016, the impact on the number of crops grown for the full sample appears to be driven by female PADMs. The impact estimate is not large but is interesting. As mentioned previously, because second season is a lesser growing season, farmers, and especially female PADMs may be more willing to try out new inputs (hybrid maize). There were also crops that were tagged, such as beans and other small crops, and that may be driving the increase.

Table 7.12a: Heterogeneous Impacts of E-verification on Secondary Outcomes by PADM Gender in First Season 2017, IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)	(5) Log maize yield in representative maize plots (tons per acre)	(6) Log value of farm production in representative maize plots (million UGX per acre)	(7) IHS* of net income of farm production in representative maize plots (million UGX)
Full Sample	0.064	-0.045	-0.017	0.286	-0.062	-0.019	-0.775*
	(0.080)	(0.069)	(0.064)	(0.426)	(0.068)	(0.120)	(0.397)
Female PADM	0.012	-0.014	-0.085	-0.164	0.096	-0.119	-1.920**
	(0.151)	(0.147)	(0.112)	(0.714)	(0.143)	(0.235)	(0.839)
Male PADM	0.025	-0.091	0.020	0.413	-0.141*	-0.050	-0.547
	(0.107)	(0.090)	(0.074)	(0.521)	(0.082)	(0.147)	(0.498)
P-value of difference in coefficients	0.957	0.771	0.603	0.611	0.315	0.860	0.303
Number of observations - Female PADM	697	533	659	665	527	541	532
Number of observations - Male PADM	1,271	1,056	1,214	1,237	1,033	1,044	1,032
Endline control group mean - Female PADM	4.161	4.683	12.418	12.296	5.718	13.144	11.567
Endline control group mean - Male PADM	4.237	5.006	12.546	12.329	5.980	13.583	12.085

Notes: IHS is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table 12b: Heterogeneous Impacts of E-verification on Secondary Outcomes by PADM Gender in Second Season 2016, IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)
Full Sample	0.185**	0.003	0.007	0.491
	(0.081)	(0.089)	(0.054)	(0.326)
Female PADM	0.265*	-0.050	0.145	0.758
	(0.158)	(0.178)	(0.118)	(0.718)
Male PADM	0.056	0.033	-0.048	0.441
	(0.109)	(0.104)	(0.062)	(0.340)
P-value of difference in coefficients	0.451	0.781	0.313	0.778
Number of observations - Female PADM	696	499	631	638
Number of observations - Male PADM	1,267	953	1,175	1,191
Endline control group mean - Female PADM	3.622	4.642	12.552	12.314
Endline control group mean - Male PADM	3.649	5.026	12.730	13.149

Notes: IHS is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

We next consider whether there are differential impacts of the e-verification intervention by the education level of the PADM. We compare impacts between households in which the PADM has completed primary school or more and households in which the PADM has not completed primary school. Table 7.13a displays results for first season 2017 and Table 7.13b displays results for second season 2016. As expected, we see that education is associated with more technology adoption: PADMs with more education use more of each input in the endline control group in each season. Turning to the impact estimates, there does not appear to be a systematic difference in the impact of e-verification based on the education level of the PADM. In first season 2017, there is a significant impact on adoption of any maize, any fertilizer and (weakly) hybrid maize adoption compared to the control group for those who did not complete primary education. This result indicates that PADMs with lower schooling levels were not disadvantaged compared to those who had completed primary school. There are no statistically significant differences between PADMs who have and have not completed primary school but given that the sample size is much smaller for those who have completed primary school or more, we may not be able to detect effects either between PADMs with primary education or more and the control group or between the two groups with differing levels of education. In second season 2016, there are no significant differences between the two groups and their respective control groups, or between each other. Once again, second season is a less intensive growing season and fewer households tend to purchase inputs at that time.

Table 7.13a: Heterogeneous Impacts of E-verification on the Use of Agricultural Inputs by Education Level in First Season 2017, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Full Sample	0.054***	0.056***	0.077***	0.066***	0.032*	0.053***
	(0.017)	(0.018)	(0.023)	(0.023)	(0.017)	(0.020)
PADM completed primary school or more	0.063	0.009	0.115**	0.098*	-0.005	0.028
	(0.059)	(0.037)	(0.056)	(0.057)	(0.058)	(0.067)
PADM did not complete primary school	0.037*	0.044**	0.085***	0.086***	0.030	0.052**
	(0.020)	(0.021)	(0.026)	(0.026)	(0.019)	(0.024)
P-value of difference in coefficients	0.777	0.594	0.775	0.904	0.693	0.792
Number of observations - PADM completed primary school or more	410	412	390	399	399	398
Number of observations - PADM did not complete primary school	1,539	1,546	1,504	1,523	1,523	1,522
Endline control group mean - PADM completed primary school or more	0.170	0.905	0.550	0.571	0.143	0.200
Endline control group mean - PADM did not complete primary school	0.090	0.887	0.413	0.428	0.091	0.139

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

In terms of secondary outcomes, it is net income again where we see differential impacts. Table 7.14a displays results for first season 2017 and Table 7.14b for second season 2016. This dimension of heterogeneity appears to matter quite a lot. In first season 2017, e-verification causes net income to increase for PADMs with higher education levels relative to high education PADMs in the control group (by a large value of 2 million UGX, or 18 percent); e-verification decreases net income for PADMs with less than primary education. Possible explanations for this pattern include that the productivity of e-verification is higher for more educated farmers because they may be verifying the e-verified products more, they may have been more likely to be offered tagged inputs by shop owners (Bold et al, 2018 show that farmer characteristics impact the quality of products they purchase), or they may be able to better to apply inputs correctly (for example, follow the instructions on the labels) and this could interact with better quality inputs as their effects will be more pronounced when applied appropriately.

Table 7.13b: Heterogeneous Impacts of E-verification on the Use of Agricultural Inputs by Education Level in Second Season 2016, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Full Sample	0.031**	0.037*	0.005	0.012	0.002	0.001
	(0.015)	(0.020)	(0.023)	(0.023)	(0.014)	(0.017)
PADM completed primary school or more	0.057	-0.019	0.035	-0.004	-0.015	0.007
	(0.040)	(0.056)	(0.056)	(0.060)	(0.042)	(0.056)
PADM did not complete primary school	0.012	0.033	0.017	0.025	-0.013	-0.014
	(0.016)	(0.025)	(0.026)	(0.026)	(0.015)	(0.017)
P-value of difference in coefficients	0.508	0.545	0.858	0.774	0.975	0.816
Number of observations - PADM completed primary school or more	400	403	375	392	392	392
Number of observations - PADM did not complete primary school	1,479	1,492	1,433	1,478	1,478	1,478
Endline control group mean - PADM completed primary school or more	0.128	0.818	0.459	0.486	0.082	0.138
Endline control group mean - PADM did not complete primary school	0.065	0.807	0.323	0.343	0.058	0.092

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

In second season 2016, the pattern of impacts changes. Those with lower levels of education grew more crops as a result of the e-verification intervention, and those with higher levels of education did not. This then translates into a higher value of farm production for those with lower education levels (possibly driven by the larger number of crops) and positive net income for this group. The difference is large (68 percent) and statistically significant. For those with higher levels of education, the lower number of crops grown (possibly to focus on traditionally grown crops – there was no impact of e-verification on take-up of more crops for this group) along with higher investments in tagged inputs may have translated into lower value of production and net income. These results are also consistent with those for input adoption. It was the less educated PADMs who were more likely to take up inputs. It is not clear however, why this pattern of impacts differ from the results for first season 2017. This issue requires further investigation.

Another important dimension for differential effects of the e-verification intervention may be poverty. We test for differential effects of the program by household poverty status, measured using an asset index. During the baseline household survey, we asked households about their ownership of assets from a list of 32 assets. We then aggregate these assets using the first component of a principal components analysis. This method ensures that a common component of the meaning of each asset measure is being captured in the same way. We separate households as being in the bottom quintile of the asset index and those who are not in the bottom quintile. The households in the bottom quintile are the poorest households in the sample. Table 7.15a presents results for first season 2017. We find a pattern of input adoption by poverty status that is consistent with patterns of other dimensions of heterogeneity found earlier. Control group households (panel 3 in Table 7.13a) in the poorest quintile are less likely to take up inputs than less poor households in general. For example, only 5.1 percent of control group households in the poorest asset quintile adopted hybrid maize first season 2017, while 12.3 percent of other households did. For the impact estimates, there is no particular pattern of differential impacts across the outcomes for take up of inputs; in some cases, the estimated impacts are higher for poor households than for the non-poor and in other cases estimated impacts on the poor are higher. The impacts of the intervention are statistically significant for take-up of all inputs for the non-poor group (compared to the non-poor control group), but there are no statistically significant differences in impacts between the poor and non-poor groups. In the poorest quintile, only impacts on hybrid maize are statistically significant; weaker effects for the other inputs sometimes reflects low power due to the small sample.

Table 7.15b presents results for second season 2016. Here, we do see a statistically significant difference ($p<0.05$) between poor and non-poor groups for the take-up of hybrid maize seed, where the impact of the e-verification intervention was higher among poorer households, and because the base control group mean was so low, impacts were huge in terms of magnitude. E-verification caused households in the poorest quintile of the asset index to increase take-up of hybrid maize seed by 15.3 percentage points from only 2.4 percentage points in the control group. The intervention induced the poorest households to try a new product. This result is of substantial interest in part because it is commonly believed that it is difficult to promote adoption of modern agricultural inputs among the poor. Interestingly, we also see that there were negative impacts of the intervention on the use of herbicide and fertilizer by the poor. Although we cannot make definitive claims about the reasons behind these impacts, it is possible that poor households were switching from using herbicide and fertilizer (which was not tagged) to hybrid maize seed. It is possible that the impacts in second season 2016 are more pronounced because this is a less important growing season so trying a new product may be less risky. The pattern of effects suggests that e-verification induced the poorest households to all but abandon fertilizer and reduce herbicide use in order to increase hybrid maize adoption. Poorer households also have the most to lose from lower quality inputs. In both first season 2017 and second season 2016, the impacts on the two groups help us understand why the impacts on the full sample are small and statistically insignificant. There are sometimes opposite effects for the two groups, and the full sample result is a weighted average of the two. It is only for hybrid maize where we see full sample and sub-sample impacts that are statistically significant. Overall, this evidence suggests that poorer households were not disadvantaged in the e-verification scheme and were just as likely to respond to it.

Table 7.14a: Heterogeneous Impacts of E-verification on Secondary Outcomes in First Season 2017, IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)	(5) Log maize yield in representative maize plots (tons per acre)	(6) Log value of farm production in representative maize plots (million UGX per acre)	(7) IHS* of net income of farm production in representative maize plots (million UGX)
Full Sample	0.064	-0.045	-0.017	0.286	-0.062	-0.019	-0.775*
	(0.080)	(0.069)	(0.064)	(0.426)	(0.068)	(0.120)	(0.397)
PADM completed primary school or more	-0.042	-0.115	-0.144	0.766	-0.070	-0.108	2.152**
	(0.169)	(0.148)	(0.128)	(1.171)	(0.123)	(0.260)	(1.013)
PADM did not complete primary school	0.071	-0.043	0.010	0.067	-0.102	0.023	-1.442***
	(0.091)	(0.080)	(0.076)	(0.420)	(0.090)	(0.155)	(0.497)
P-value of difference in coefficients	0.699	0.798	0.511	0.674	0.898	0.768	0.039
Number of observations - PADM completed primary school or more	413	342	385	394	344	349	340
Number of observations - PADM did not complete primary school	1,555	1,247	1,488	1,508	1,216	1,236	1,224
Endline control group mean - PADM completed primary school or more	4.436	5.087	12.696	12.076	6.077	13.708	11.668
Endline control group mean - PADM did not complete primary school	4.153	4.854	12.454	12.381	5.847	13.368	11.988

Notes: IHS is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table 7.14b: Heterogeneous Impacts of E-verification on Secondary Outcomes in Second Season 2016, IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)
Full Sample	0.185**	0.003	0.007	0.491
	(0.081)	(0.089)	(0.054)	(0.326)
PADM completed primary school or more	-0.367*	-0.195	-0.402***	0.318
	(0.203)	(0.224)	(0.111)	(0.572)
PADM did not complete primary school	0.259***	0.037	0.134**	0.877**
	(0.097)	(0.104)	(0.065)	(0.399)
P-value of difference in coefficients	0.054	0.548	0.016	0.580
Number of observations - PADM completed primary school or more	411	305	382	386
Number of observations - PADM did not complete primary school	1,552	1,147	1,424	1,443
Endline control group mean - PADM completed primary school or more	3.961	5.033	12.840	13.737
Endline control group mean - PADM did not complete primary school	3.556	4.873	12.626	12.650

Notes: IHS is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * $p < 0.1$, ** $p < .05$, *** $p < 0.01$.

Table 7.15a: Heterogeneous Impacts of E-verification on the Use of Agricultural Inputs by Poverty Status in First Season 2017, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Full Sample	0.054***	0.056***	0.077***	0.066***	0.032*	0.053***
	(0.017)	(0.018)	(0.023)	(0.023)	(0.017)	(0.020)
Lowest quintile of all asset index	0.097***	0.037	0.103	0.084	-0.055*	0.021
	(0.036)	(0.058)	(0.067)	(0.069)	(0.033)	(0.039)
Not in the lowest quintile of all asset index	0.049**	0.048**	0.076***	0.074***	0.036*	0.049**
	(0.021)	(0.019)	(0.028)	(0.027)	(0.020)	(0.023)
P-value of difference in coefficients	0.443	0.901	0.811	0.923	0.159	0.714
Number of observations - Lowest quintile of all asset index	374	376	366	367	367	366
Number of observations - Not in the lowest quintile of all asset index	1,535	1,540	1,488	1,514	1,514	1,513
Endline control group mean - Lowest quintile of all asset index	0.051	0.849	0.383	0.396	0.056	0.057
Endline control group mean - Not in the lowest quintile of all asset index	0.123	0.898	0.451	0.468	0.114	0.176

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table 7.15b: Heterogeneous Impacts of E-verification on the Use of Agricultural Inputs by Poverty Status in Second Season 2016 with IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Full Sample	0.031**	0.037*	0.005	0.012	0.002	0.001
	(0.015)	(0.020)	(0.023)	(0.023)	(0.014)	(0.017)
Lowest quintile of all asset index	0.153***	0.102	-0.066	-0.128*	-0.052	-0.092**
	(0.037)	(0.073)	(0.061)	(0.066)	(0.035)	(0.047)
Not in the lowest quintile of all asset index	0.009	0.018	0.021	0.034	0.015	0.015
	(0.017)	(0.022)	(0.027)	(0.026)	(0.015)	(0.018)
P-value of difference in coefficients	0.028	0.461	0.438	0.182	0.277	0.143
Number of observations - Lowest quintile of all asset index	356	359	344	351	351	351
Number of observations - Not in the lowest quintile of all asset index	1,484	1,496	1,428	1,480	1,480	1,480
Endline control group mean - Lowest quintile of all asset index	0.024	0.782	0.290	0.320	0.054	0.080
Endline control group mean - Not in the lowest quintile of all asset index	0.091	0.817	0.365	0.384	0.066	0.107

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

We next turn to impacts of secondary outcomes by poverty status. Table 7.16a shows impact estimates for first season 2017, and Table 7.16b for second season 2016. For farmers considered poor, a consistent result is that the e-verification intervention generally had negative impacts, although most are not statistically significant. We saw in the two preceding tables that in first season 2017 and in second season 2016, poorer households appeared to be substituting away from fertilizer and herbicide and towards trying hybrid maize seeds. This would explain the negative coefficients on number of crops, total yields if they only tried hybrid maize on a small area, and thus net income. For the non-poor group, impact estimates are almost always positive, though only the impact on net income in second season 2016 is significant. Table 7.19b shows that it was the non-poor households who were trying new crops and had higher net income due to the e-verification intervention, and that this is what is driving the results in the full sample. There is likely substantial overlap between education levels and poverty. Thus, although e-verification prompted significant new input adoption by poor households, the effects on their productivity and income were negative.

Table 7.16a: Heterogeneous Impacts of E-verification of Poverty on Secondary Outcomes in First Season 2017 with IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)	(5) Log maize yield in representative maize plots (tons per acre)	(6) Log value of farm production in representative maize plots (million UGX per acre)	(7) IHS* of net income of farm production in representative maize plots (million UGX)
Full Sample	0.064 (0.080)	-0.045 (0.069)	-0.017 (0.064)	0.286 (0.426)	-0.062 (0.068)	-0.019 (0.120)	-0.775* (0.397)
Lowest quintile of asset index	-0.188 (0.204)	-0.564** (0.270)	-0.476** (0.194)	-0.307 (0.856)	-0.077 (0.250)	-0.252 (0.410)	-1.508 (1.561)
Not in the lowest quintile of asset index	0.069 (0.092)	0.047 (0.077)	0.011 (0.072)	0.474 (0.522)	-0.103 (0.073)	0.046 (0.138)	-0.357 (0.447)
P-value of difference in coefficients	0.452	0.160	0.109	0.610	0.942	0.633	0.630
Number of observations - Lowest quintile of asset index	379	289	361	361	286	291	286
Number of observations - Not in the lowest quintile of asset index	1,547	1,265	1,473	1,502	1,239	1,259	1,243
Endline control group mean - Lowest quintile of asset index	3.800	5.038	12.498	12.353	5.856	13.495	12.128
Endline control group mean - Not in the lowest quintile of asset index	4.295	4.882	12.506	12.331	5.903	13.404	11.882

Notes: IHS is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table 7.16b: Heterogeneous Impacts of E-verification on Secondary Outcomes in Second Season 2016, IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)
Full Sample	0.185**	0.003	0.007	0.491
	(0.081)	(0.089)	(0.054)	(0.326)
Lowest quintile of asset index	0.077	-0.218	-0.207	-0.075
	(0.251)	(0.297)	(0.161)	(1.087)
Not in the lowest quintile of asset index	0.152*	0.066	-0.004	0.691**
	(0.088)	(0.095)	(0.064)	(0.342)
P-value of difference in coefficients	0.851	0.565	0.505	0.647
Number of observations - Lowest quintile of asset index	379	269	336	339
Number of observations - Not in the lowest quintile of asset index	1,542	1,158	1,433	1,452
Endline control group mean - Lowest quintile of asset index	3.196	5.003	12.636	12.373
Endline control group mean - Not in the lowest quintile of asset index	3.746	4.871	12.676	13.039

Notes: IHS is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

We next turn to differences between PADMs based on preferences over risk and ambiguity. Table 7.17a displays results for PADMs who are relatively more risk-averse compared to those who are relatively less risk averse for first season 2017 and Table 7.17b for second season 2016. Risk-aversion is measured using a household survey question that asks respondents to choose between two seeds: a riskier seed that has high payouts when rains are enough but low payouts when rains are little, and a less risky seed that has a medium level of payoff no matter what. The chance of enough and little rain is 50 percent. With this framing, households that choose the riskier seed are *relatively* less risk averse. This is why we use the terms more risk-averse and less risk-averse; identifying risk-averse and not risk-averse in an absolute sense would require more complicated framing in which outcomes for each seed changed with rain probabilities. We see that in the control group, the probability of adoption for each input does not vary much with the relative risk aversion of the PADM. In addition, estimated impacts of e-verification on input adoption are quite similar for the more- and less-risk-averse households. The percentage increase in adoption for both groups is largest for hybrid maize (30 percent for relatively more risk-averse and 80 percent for those less risk-averse). We see that the response to e-verification by less risk-averse household is to increase adoption of hybrid maize by 7.6 percent, while the impact on adoption of hybrid maize for more risk averse households is 3.7 percent (the difference is not significant). Part of this difference in estimated impacts on hybrid maize adoption may be due to differences in the control group means. As with the full sample results the impacts are generally positive for both groups and in first season statistically significantly different from the control group in some cases in first season 2017. In second season 2016 the results again are insignificant, as are the full sample results.

Table 7.17a: Heterogeneous Impacts of E-verification on the Use of Agricultural Inputs by Risk Preferences in First Season 2017, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Full Sample	0.054***	0.056***	0.077***	0.066***	0.032*	0.053***
	(0.017)	(0.018)	(0.023)	(0.023)	(0.017)	(0.020)
More risk averse	0.037*	0.045**	0.089***	0.083***	0.024	0.050**
	(0.021)	(0.022)	(0.028)	(0.028)	(0.022)	(0.025)
Less risk averse	0.076**	0.031	0.087**	0.083**	0.054	0.055
	(0.034)	(0.033)	(0.040)	(0.038)	(0.033)	(0.039)
P-value of difference in coefficients	0.508	0.801	0.979	0.997	0.593	0.932
Number of observations - More risk averse	1,368	1,378	1,336	1,350	1,350	1,349
Number of observations - Less risk averse	579	578	556	570	570	569
Endline control group mean - More risk averse	0.112	0.892	0.437	0.454	0.109	0.160
Endline control group mean - Less risk averse	0.095	0.887	0.450	0.466	0.086	0.133

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table 7.17b: Heterogeneous Impacts of E-verification on the Use of Agricultural Inputs by Risk Preferences in Second Season 2016, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Full Sample	0.031**	0.037*	0.005	0.012	0.002	0.001
	(0.015)	(0.020)	(0.023)	(0.023)	(0.014)	(0.017)
More risk averse	0.019	0.016	0.017	0.021	0.001	0.001
	(0.018)	(0.027)	(0.030)	(0.030)	(0.016)	(0.020)
Less risk averse	0.022	0.060	-0.008	-0.009	-0.018	-0.020
	(0.028)	(0.042)	(0.041)	(0.040)	(0.027)	(0.035)
P-value of difference in coefficients	0.945	0.563	0.752	0.713	0.668	0.698
Number of observations - More risk averse	1,321	1,333	1,269	1,315	1,315	1,315
Number of observations - Less risk averse	556	560	538	553	553	553
Endline control group mean - More risk averse	0.082	0.829	0.334	0.359	0.065	0.104
Endline control group mean - Less risk averse	0.070	0.764	0.393	0.406	0.060	0.096

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table 7.18a presents results for first season 2017, and Table 7.18b for second season 2016 for secondary outcomes. In first season 2017, there are no statistically significant differences in outcomes for either of the two groups compared to their control groups. There are also no statistically significant differences between the two groups. As with take-up of inputs, we also see that the control group means are not very different. For second season 2016, we see that it is the more risk-averse farmers who tend to try new crops. For these outcomes, control group means are quite similar, so the impact estimates are not being driven by these. There may also be a correlation between risk-aversion, education, and poverty.

Table 7.18a: Heterogeneous Impacts of E-verification on Secondary Outcomes by Risk Preferences in First Season 2017, IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)	(5) Log maize yield in representative maize plots (tons per acre)	(6) Log value of farm production in representative maize plots (million UGX per acre)	(7) IHS* of net income of farm production in representative maize plots (million UGX)
Full Sample	0.064	-0.045	-0.017	0.286	-0.062	-0.019	-0.775*
	(0.080)	(0.069)	(0.064)	(0.426)	(0.068)	(0.120)	(0.397)
More risk averse	-0.013	-0.039	-0.028	-0.439	-0.004	-0.034	-0.593
	(0.094)	(0.078)	(0.076)	(0.476)	(0.077)	(0.151)	(0.528)
Less risk averse	0.153	-0.104	-0.202*	1.183	-0.079	0.094	0.144
	(0.153)	(0.127)	(0.108)	(0.735)	(0.126)	(0.253)	(0.886)

P-value of difference in coefficients	0.509	0.775	0.342	0.186	0.740	0.750	0.598
Number of observations - More risk averse	1,383	1,116	1,310	1,332	1,098	1,117	1,100
Number of observations - Less risk averse	583	471	561	568	460	466	462
Endline control group mean - More risk averse	4.235	4.918	12.525	12.521	5.908	13.440	11.825
Endline control group mean - Less risk averse	4.155	4.871	12.456	11.862	5.873	13.430	12.112

Notes: IHS is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table 7.18b: Heterogeneous Impacts of E-verification on Secondary Outcomes by Risk Preferences in Second Season 2016 with IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)
Full Sample	0.185**	0.003	0.007	0.491
	(0.081)	(0.089)	(0.054)	(0.326)
More risk averse	0.235**	-0.111	-0.061	0.375
	(0.102)	(0.100)	(0.064)	(0.379)
Less risk averse	-0.024	-0.008	0.036	0.908
	(0.168)	(0.185)	(0.100)	(0.620)
P-value of difference in coefficients	0.350	0.745	0.610	0.599
Number of observations - More risk averse	1,378	1,042	1,268	1,285
Number of observations - Less risk averse	583	408	536	542
Endline control group mean - More risk averse	3.633	4.927	12.689	12.928
Endline control group mean - Less risk averse	3.651	4.857	12.631	12.760

Notes: IHS is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

We now examine how impacts vary with ambiguity aversion. This preference was measured using a question in the household survey asking respondents to strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree with the following statement: "I feel uncomfortable in situations in which I do not know the likelihood of different outcomes". Respondents are classified as ambiguity-averse if they respond with strongly agree or agree. Table 7.19a reports results for the two groups for first season 2017 and Table 7.19b for second season 2016. The impacts of e-verification are consistently positive and often statistically significant for the sample who are not ambiguity-averse, and this is true both for first season 2017 and for second season 2016. These results are consistent with the impact estimates for the full sample. The coefficients are often larger for the group that is not ambiguity-averse, but we cannot reject that the two groups have statistically similar impact estimates. The sample size for the ambiguity-averse group is small, however, so caution is needed when interpreting these results. For second season 2016, we do see negative and significant impacts on herbicide adoption for the ambiguity-averse group and positive and significant coefficients for the group who is not ambiguity-averse. This pattern is unexpected – we might have expected that e-verified products may have been more attractive to farmers who were staying out of the market due to ambiguity aversion over the quality of herbicides in the market. Our results show that it was households that were not ambiguity averse who responded more to e-verification. This suggests that households may have believed that the benefits of the e-verification were ambiguous themselves. These estimates help explain why there is not a significant effect for the full sample. It is possible that herbicide had much more ambiguity regarding authenticity compared to the other inputs and that e-verification reduced this uncertainty, leading to increased take-up for ambiguity-averse PADMs in herbicide in second season (where inputs in shops may be left over from first season) but not as much for the other inputs, including fertilizer, which was not tagged. As with risk-aversion, take-up of glyphosate herbicide and any herbicide is higher among ambiguity-averse households than non-ambiguity averse households in the control group. This suggests that the treatment effect is due in part to the differences in control group means. Additionally, it is possible that, because e-verification began in second season 2016, it took more ambiguity-averse farmers time to learn the benefits of the new inputs since they are less likely to try new inputs. They may have seen the effects from neighbors or friends and then decided to adopt. Although not statistically significant, the coefficients are positive for the ambiguity-averse group in first season 2017.

Table 7.19a: Heterogeneous Impacts of E-verification on the Use of Agricultural Inputs by Ambiguity Aversion in First Season 2017, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Full Sample	0.054***	0.056***	0.077***	0.066***	0.032*	0.053***
	(0.017)	(0.018)	(0.023)	(0.023)	(0.017)	(0.020)
Ambiguity averse	0.049	0.033	0.032	0.019	0.006	0.031
	(0.031)	(0.033)	(0.044)	(0.043)	(0.036)	(0.040)
Not ambiguity averse	0.057**	0.057***	0.111***	0.107***	0.028	0.042
	(0.022)	(0.021)	(0.031)	(0.032)	(0.023)	(0.026)
P-value of difference in coefficients	0.892	0.661	0.272	0.239	0.697	0.869
Number of observations - Ambiguity averse	697	697	677	684	684	682
Number of observations - Not ambiguity averse	1,249	1,258	1,214	1,235	1,235	1,235
Endline control group mean - Ambiguity averse	0.130	0.893	0.490	0.500	0.109	0.140
Endline control group mean - Not ambiguity averse	0.094	0.889	0.413	0.433	0.098	0.158

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * $p < 0.1$, ** $p < .05$, *** $p < 0.01$.

Table 7.19b: Heterogeneous Impacts of E-verification on the Use of Agricultural Inputs by Ambiguity Aversion in Second Season 2016, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Full Sample	0.031**	0.037*	0.005	0.012	0.002	0.001
	(0.015)	(0.020)	(0.023)	(0.023)	(0.014)	(0.017)
Ambiguity averse	0.027	0.027	-0.104**	-0.108**	-0.014	0.006
	(0.031)	(0.038)	(0.045)	(0.045)	(0.022)	(0.027)
Not ambiguity averse	0.030*	0.030	0.060**	0.064**	0.010	-0.009
	(0.018)	(0.029)	(0.029)	(0.029)	(0.019)	(0.021)
P-value of difference in coefficients	0.958	0.962	0.031	0.021	0.588	0.780
Number of observations - Ambiguity averse	673	679	640	668	668	668
Number of observations - Less ambiguity averse	1,203	1,213	1,166	1,199	1,199	1,199
Endline control group mean - Ambiguity averse	0.091	0.813	0.417	0.444	0.079	0.099
Endline control group mean - Not ambiguity averse	0.071	0.806	0.316	0.335	0.055	0.103

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

In terms of ambiguity-aversion, the patterns for the secondary outcomes are again consistent with impact estimates on input adoption. Tables 7.20a and 7.20b display results for first season 2017 and second season 2016, respectively. PADMs who are not ambiguity-averse were more likely to take up inputs, but we see no pattern of significant impacts on yields and net income for these households. For ambiguity-averse PADMs, impact estimates for the two measures of yield in first season 2017 show a negative effect of e-verification on yields. However, these effects do not appear to be driven by changes in take up of e-verified products (shown earlier). In second season 2016, ambiguity-averse PADMs experience a small positive impact on net income.

Table 7.20a: Heterogeneous Impacts of E-verification on Secondary Outcomes by Ambiguity Aversion in First Season 2017, IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)	(5) Log maize yield in representative maize plots (tons per acre)	(6) Log value of farm production in representative maize plots (million UGX per acre)	(7) IHS* of net income of farm production in representative maize plots (million UGX)
Full Sample	0.064	-0.045	-0.017	0.286	-0.062	-0.019	-0.775*
	(0.080)	(0.069)	(0.064)	(0.426)	(0.068)	(0.120)	(0.397)
Ambiguity averse	-0.014	-0.331***	0.017	0.303	-0.191	-0.560**	-0.716
	(0.127)	(0.118)	(0.122)	(0.732)	(0.118)	(0.221)	(0.683)
Not ambiguity averse	0.007	0.127	-0.089	0.240	-0.014	0.174	-0.636
	(0.096)	(0.090)	(0.072)	(0.485)	(0.085)	(0.150)	(0.506)
P-value of difference in coefficients	0.925	0.027	0.567	0.959	0.388	0.050	0.951

Number of observations - Ambiguity averse	700	571	667	679	552	561	554
Number of observations - Not ambiguity averse	1,265	1,015	1,203	1,220	1,005	1,021	1,007
Endline control group mean - Ambiguity averse	4.192	5.058	12.522	12.200	5.977	13.738	12.467
Endline control group mean - Not ambiguity averse	4.218	4.817	12.494	12.381	5.856	13.279	11.614

Notes: IHS is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table 7.20b: Heterogeneous Impacts of E-verification on Secondary Outcomes by Ambiguity Aversion in Second Season 2016, IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)
Full Sample	0.185**	0.003	0.007	0.491
	(0.081)	(0.089)	(0.054)	(0.326)
Ambiguity averse	0.116	0.179	0.139	0.990*
	(0.134)	(0.144)	(0.094)	(0.548)
Not ambiguity averse	0.114	-0.025	-0.081	0.475
	(0.104)	(0.123)	(0.072)	(0.419)
P-value of difference in coefficients	0.992	0.423	0.217	0.590
Number of observations - Ambiguity averse	697	532	656	660
Number of observations - Not ambiguity averse	1,263	917	1,147	1,166
Endline control group mean - Ambiguity averse	3.654	5.034	12.686	12.876
Endline control group mean - Not ambiguity averse	3.628	4.835	12.664	12.875

Notes: IHS is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Finally, we examine heterogeneous impacts of the e-verification intervention on households by phone ownership. PADMs who own or have access to a phone were more likely to be exposed to the encouragement promotion that operated through the IVR messages and possibly through the promotion meetings. Those without a phone cannot access these messages unless someone else shares them. All PADMs, regardless of phone access, could attend the community promotion meetings; however, it is possible that if community members were being gathered by phone, they may also have had lower attendance at these meetings. Also, it would have been easier for phone owners to verify that their e-tag- or Ag verify-labeled inputs are genuine. The non-phone owning group was purposely included in the sample to see if there may be spillover effects of the encouragement intervention onto non-phone owners. This would tell us whether such an intervention would necessarily need to reach all households in an LC1 or could reach fewer as there may be spillover effects. In this case, because of the CPMs, the spillovers may be muted, but it is still interesting to compare the patterns.

Table 7.21a reports impact estimates for the sample of PADMs who own a phone or have access to one for first season 2017, and Table 7.21b for second season 2016. A first pattern is that, as expected, in the control group PADMs with phones tend to use more inputs. This characteristic is likely to be correlated with characteristics such as asset ownership, education, and income. In first season 2017, both the phone-owning and non-phone owning groups have positive coefficients. The impacts for the phone-owning group are positive and statistically significant compared to the phone-owning control group and are larger than the impacts for the non-phone owning group in most cases. The impacts for the non-phone owning group are also always positive and are larger than those of the phone-owning group for non-tagged products (any maize – with a statistically significant difference between the two groups, p-value 0.029 – and any fertilizer) compared to hybrid maize and glyphosate herbicide. This result suggests that there may have been an increase in input adoption in general through spillovers or via the CPMs. In second season 2016, the same pattern holds for any maize, but not for the other inputs. In that season, impacts for both groups are generally statistically insignificant. The negative coefficients are very small in magnitude and their standard errors are fairly large. Taken together, these results suggest that there may have been some spillovers in adoption of non-tagged inputs to non-phone owners, but non-phone owners may also have been exposed to encouragement promotions through the CPMs. If so, these impacts represent experimentally induced direct effects rather than spillovers.

Table 7.21a: Heterogeneous Impacts of E-verification on the Use of Agricultural Inputs by Phone Access in First Season 2017, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Full Sample	0.054***	0.056***	0.077***	0.066***	0.032*	0.053***
	(0.017)	(0.018)	(0.023)	(0.023)	(0.017)	(0.020)
Has phone	0.056***	0.037**	0.095***	0.084***	0.033	0.043*
	(0.021)	(0.018)	(0.026)	(0.025)	(0.020)	(0.023)
No phone	0.027	0.183***	0.027	0.008	0.018	0.087***
	(0.031)	(0.049)	(0.051)	(0.051)	(0.027)	(0.032)
P-value of difference in coefficients	0.595	0.029	0.438	0.371	0.766	0.462
Number of observations - Has phone	1,474	1,471	1,424	1,456	1,456	1,446
Number of observations - No phone	550	552	530	537	537	535
Endline control group mean - Has phone	0.118	0.904	0.479	0.493	0.120	0.177
Endline control group mean - No phone	0.065	0.840	0.310	0.327	0.045	0.065

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows:

* p < 0.1, ** p < .05, *** p < 0.01.

Table 7.21b: Heterogeneous Impacts of E-verification on the Use of Agricultural Inputs by Phone Access in Second Season 2016, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Full Sample	0.031**	0.037*	0.005	0.012	0.002	0.001
	(0.015)	(0.020)	(0.023)	(0.023)	(0.014)	(0.017)
Has phone	0.030	0.019	0.019	0.024	0.008	0.007
	(0.018)	(0.022)	(0.026)	(0.026)	(0.017)	(0.020)
No phone	0.010	0.150***	-0.014	-0.009	-0.017	-0.014
	(0.025)	(0.057)	(0.052)	(0.052)	(0.019)	(0.027)
P-value of difference in coefficients	0.685	0.120	0.701	0.698	0.511	0.691
Number of observations - Has phone	1,416	1,429	1,368	1,415	1,415	1,415
Number of observations - No phone	526	530	500	515	515	515
Endline control group mean - Has phone	0.087	0.822	0.385	0.406	0.073	0.114
Endline control group mean - No phone	0.041	0.762	0.240	0.260	0.028	0.053

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Tables 7.22a and 7.22b display impact estimates for secondary outcomes for PADMs differentiated by access to phones. Again, this dimension of heterogeneity appears to be somewhat correlated with those such as income and education. Those without phones grow more crops in both first season 2017 and second season 2016, however, counter to the results for adoption, they have lower net income (approximately 2.5 million UGX in first season 2017) as a result of e-verification. Again, it could be that these households do benefit from spillovers but are likely to be poorer and so are not able to use inputs as effectively and thus may have lower net income. While non-phone owners were not disadvantaged in benefitting from the e-verification intervention in terms of take-up of inputs, they may have been disadvantaged a bit in terms of other farm outcomes. In second season 2016, we see that non-phone owning households had positive net incomes, but that the impact is imprecisely estimated.

Table 7.22a: Heterogeneous Impacts of E-verification on Secondary Outcomes by Phone Access in First Season 2017, IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)	(5) Log maize yield in representative maize plots (tons per acre)	(6) Log value of farm production in representative maize plots (million UGX per acre)	(7) IHS* of net income of farm production in representative maize plots (million UGX)
Full Sample	0.064	-0.045	-0.017	0.286	-0.062	-0.019	-0.775*
	(0.080)	(0.069)	(0.064)	(0.426)	(0.068)	(0.120)	(0.397)
Has phone	-0.038	-0.082	-0.012	0.107	-0.102	-0.006	-0.529
	(0.087)	(0.074)	(0.067)	(0.489)	(0.073)	(0.132)	(0.456)
No phone	0.554***	-0.110	-0.012	0.856	0.088	-0.424	-2.561***
	(0.170)	(0.175)	(0.133)	(0.603)	(0.157)	(0.266)	(0.926)
P-value of difference in coefficients	0.026	0.926	0.996	0.547	0.472	0.342	0.174
Number of observations - Has phone	1,478	1,218	1,411	1,436	1,196	1,218	1,206
Number of observations - No phone	557	423	521	526	415	420	410
Endline control group mean - Has phone	4.305	4.943	12.525	12.375	5.939	13.440	11.766
Endline control group mean - No phone	3.872	4.772	12.405	11.832	5.725	13.311	12.359

Notes: IHS is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table 7.22b: Heterogeneous Impacts of E-verification on Secondary Outcomes by Ambiguity Aversion in Second Season 2016, IV Specifications

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income of farm production (million UGX)
Full Sample	0.185**	0.003	0.007	0.491
	(0.081)	(0.089)	(0.054)	(0.326)
Has phone	0.048	0.145	0.001	0.439
	(0.089)	(0.096)	(0.061)	(0.373)
No phone	0.580***	-0.500**	0.012	0.301
	(0.189)	(0.215)	(0.122)	(0.502)
P-value of difference in coefficients	0.083	0.080	0.961	0.896
Number of observations - Has phone	1,474	1,107	1,362	1,383
Number of observations - No phone	556	391	502	505
Endline control group mean - Has phone	3.719	4.858	12.673	12.912
Endline control group mean - No phone	3.350	5.070	12.662	12.863

Notes: IHS is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. The outcome variables in the regressions in columns (2)-(7) are trimmed at 3 standard deviations above and below the mean before taking log values. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

The results in this section indicate that the e-verification intervention did have differential impacts on different types of households. For gender, male PADMs experienced higher take-up of inputs as well as net income. PADMs with less education experienced bigger impacts from e-verification on input adoption, but in first season 2017, had lower net income. This may be because they were less able to use the inputs once purchased. Poorer households adopted hybrid maize and substituted away from fertilizer and herbicide, particularly in second season 2016 since they had to choose how to spend their income and could perhaps try more things in the lesser growing season. The impact of e-verification did not differ much based on preferences for risk-aversion in terms of adoption or farm outcomes in either season. In second season 2016, households that were ambiguity-averse were less likely to try herbicide, but in first season impacts for both groups are positive. It is possible that ambiguity-averse households needed more time to learn about new products and see others using them before they themselves tried them. Households with phones had higher impacts in terms of take up of tagged products and households without phones had higher impacts in terms of take up of non-tagged products. This suggests some small potential spillovers as well. An interesting pattern emerges when we look at the number of crops grown and net income in second season 2016. For the full sample, the e-verification intervention leads to more crops being grown, but a negative (but statistically insignificant) impact on net income. Differential effects by groups help explain this pattern. PADMs with lower levels of education, poorer households, and more risk- and ambiguity-averse households have large impacts of e-verification on growing more crops. They also generally have higher net incomes, which are likely driven by the adoption of new crops (some of which were tagged). In their counterpart groups, the impact is sometimes less pronounced and sometimes leads to negative net incomes. Because the other group is generally the larger group in terms of sample size, the overall effect on net income turns out to be negative. Taken together, these results show that impacts of interventions may well have different and unexpected impacts on sub-groups, and that an understanding of these differential impacts can help in explaining overall impacts.

8. Impacts on Counterfeiting

8.1 Design of Counterfeiting Sub-Study

In this chapter, we will examine the third objective of this study; a reduction in the prevalence of poor-quality agricultural inputs in the market as a result of the e-verification scheme. The goals from the counterfeiting sub-study were two-fold: first, to obtain improved estimates of the quality of inputs in the market given such estimates are very rare, and second, to estimate whether e-verification's product assurance could reduce the prevalence of low-quality agricultural inputs. Recall from chapter 3 that a different research design was needed to estimate the impacts of the e-verification scheme on input quality. Since agricultural products are sold in markets ("market locations" in our study jargon), random variation in the intensity of the e-verification scheme needed to be introduced at the market level in order to test the effect of the scheme on input quality. Experimentally induced differential take-up at the market location level would allow us to identify impacts. In 60 of the 120 market locations, we randomly assigned the encouragement LC1 to receive discount vouchers for hybrid maize and glyphosate herbicide. The theory of change is that encouragement communities who received the discount would have even higher take-up of e-verified inputs compared to treatment communities that did not receive discounts. We can then compare market locations in which an encouragement LC1 did receive the discounts to market locations in which the encouragement community did not receive the discounts to estimate causal impacts of the presence of the e-verified products on overall input quality. In particular, the study was designed to test whether e-verified products themselves improved in quality and whether the presence of e-verified projects led to improvements in average quality of other inputs in the market. If the e-verification labels were seen by consumers as a signal that tagged products were less likely to be counterfeited, and demand for tagged products increased, this could reduce demand for untagged products, forcing lower quality products out of the market.

During the second round of CPMs in first season 2017, discount vouchers were handed out to all attendees at the end of the meeting.[15] Meeting participants were told that each person would receive two discount vouchers: one for purchasing Kakasa herbicide and one for purchasing Kakasa hybrid maize. The voucher for Kakasa herbicide covered the entire cost of a one-liter bottle, making it free, and the voucher for 1 kg of hybrid maize seed covered 80 percent of the cost on average. These were thus very large discounts. As explained in chapter 3, several measures were undertaken to ensure that the discount vouchers could not be tampered with (including the use of special paper not available in Uganda and careful tracking of the serial numbers of the vouchers). Three weeks (on average) after the CPM in which vouchers were handed out, the enumeration teams returned to the LC1s to provide the money for the discounts in the form of a rebate. All community members received an SMS that the enumeration team would be providing the rebates on their next visit.

[15] These vouchers were implemented as rebates, in which a voucher holder would receive a cash payment equal to the value of the voucher if they could provide the packaging for the product at a later meeting as proof that they had purchased it.

Community leaders were contacted before the visits so that they could inform and gather community members and organize a meeting space. On the day of the visit, the morning was spent further gathering community members to ensure maximum attendance.

Adherence to the study design was high: almost nobody in the control group received the discount vouchers (a few members from neighboring communities may have attended the meetings), 3 percent of people in the encouragement group in villages not assigned to the discount group received the vouchers, and about 30 percent of people in discount communities received the maize and herbicide vouchers. Based on data collected on the voucher program during the CPMs, there was a modest use of the vouchers by those that received them. On average, 55 people per community assigned to the discount treatment received a voucher for herbicide and a voucher for hybrid maize. In these communities, on average 17 people redeemed the herbicide voucher for a rebate and 10 people redeemed the hybrid maize voucher for a rebate. Voucher redemption rates may have been higher if the vouchers had been available throughout the period leading up to that season. Some farmers may have already purchased these inputs for the first growing season of 2017 before the vouchers were available. Trust that the rebates would be received, however, was likely quite high as the same enumeration team had visited the communities five times already.

Unfortunately, as has been found with other similar authentication systems, very few individuals authenticated their E-tag herbicide or hybrid maize. For herbicide, only 15 percent of people in the discount group attempted to authenticate. In the discount group, enumerators often helped with authentication during the time when the discount vouchers were redeemed. We see similarly low rates of success in authenticating E-tag maize across groups, 14 percent of the discount group was successfully able to authenticate their E-tag maize.

Starting with hybrid maize, we estimate the impact of e-verification on input quality using the following equation:

$$q_i = \pi_o + \gamma_1 D_m + d_r + \varepsilon_{im} \qquad (4)$$

where q_i is the quality measure for sample *i* at endline. D_m is a dummy variable for a market location *m* having its encouragement LC1 been assigned as a discount village, d_r is the genetic distance *d* (explained in further detail below) between two reference sample (*r*) seeds, and ε_{im} is an error term that is clustered at the level of the market location. The outcome variable is the average genetic distance from the field sample to the reference samples of the variety that the field sample claims to be. Genetic distance will be explained in greater detail below.

For herbicide, we estimate a model similar to equation (4), but without the control variable *d* for similarity of the reference samples. We measure glyphosate herbicide quality, q_i, as the ratio of measured concentration of glyphosate in the samples over the stated concentration of glyphosate on the label on the bottle.

8.2 Hybrid Maize Quality

We begin with results regarding hybrid maize seed. We first outline the method used to collect the maize seed samples and the number of samples collected, then discuss the way in which samples were tested and the number of samples tested, and finally, the way in which the data was aggregated and analyzed.

We collected maize seed samples three times: just after the baseline survey in August 2014, representing second season 2014; concurrently to the CLE2 survey in March 2015, representing first season 2015; and just before the endline survey in March 2017, representing first season 2017. The goal was to obtain a representative sample of the hybrid maize seed available in the markets visited by the farmers in our sample. If asked, enumerators collecting the maize seed samples told shop-keepers that they were students who needed to purchase some inputs for a project. They were asked not to mention anything about input quality and to ensure the owner did not think they were from the government.

For the first two rounds of data collection, we began with a randomly ordered list of all the shops in the market location that were identified in the shop listing exercise for each round. Shops were randomly ordered. Prior to sample collection, we identified the top 12 varieties of maize in terms of market share in each market hub – this was determined from the shop survey that collected information on each of the varieties sold and the quantities of each during the shop listing exercise. The enumeration team aimed to purchase 8 samples of hybrid maize from each market location, and to purchase varieties in the shops that were highest on the list. The instructions were to purchase four samples each from the first two listed shops, and if one or both of these shops had fewer than four varieties, enumerators were to go to the next shop on the list and continue in this manner until 8 samples had been collected or there were no shops left in the market. No variety was to be purchased more than once from the same shop.

Once shops with samples were identified, to select a variety, the enumerator was to count the number of different varieties available in the shop and use a random number table to select a variety. Then, a random number table was used to select the type of package of that variety that was to be purchased out of three different types (if the shop had different types of packages). The first type was an open bulk container, which is a large container or bag that is not sealed, often from a large bag (e.g. 10 kg) that was purchased by the shop. Enumerators were to have the shopkeeper measure out half a kg for purchase. The second package type was a kavera package, which is a small, sealed package that was prepared from a large bag purchased by the shop. A random number table was used to select the size of the kavera package if the shop stocked more than one size. The final packaging type was a pre-sealed smaller package that had never been opened, and enumerators aimed to purchase a 2 kg package if available, and a 5 kg package if not. Once the package type was selected, a random number table was used to select the particular package to purchase, based on the number of packages of that variety and type available in the shop. For each sample purchased, enumerators recorded, on paper, the shop and variety, the package type, the date the package was opened/purchased, the date of expiry, and the sample/package size and price. Samples collected were labeled with sample ID stickers pre-prepared by the research team and a duplicate sticker was included on the paper form. These data were later entered into a CAPI program for analysis. Once the samples were brought back to Kampala, 20 seeds from each sample were randomly selected to be tested. To select 20 seeds, the entire sample was poured onto a table that was divided into four quadrants using tape. The seeds were then mixed across the quadrants and a protocol was followed to attempt a random draw of seeds from each quadrant in order to provide a sample of 20 seeds.

For the endline round of sample collection, due to budget constraints and increased testing costs,[16] only two samples of hybrid maize seed were targeted to be purchased from each market location. One sample was to be Kakasa or Ag-verify tagged maize, and the other was to be a package of non-tagged maize. Shops from an updated shop listing exercise were again ordered randomly within market locations and were to be visited in the specified order. To select a variety, the enumerator was to count the number of different varieties (within the tagged and non-tagged categories) and use a random number table to select a variety. Shops were to be visited in order until one sample of tagged seed and one sample of non-tagged seed was purchased. Once a variety was selected, a random number table was again used to select the package type from the same three types listed above, the package size as listed above, and the specific package. Data on the samples collected were recorded in the same manner as well.

For both the first two rounds and for the third round, 'reference' hybrid maize seed samples were obtained from breeders where possible, or from the manufacturers. These would serve as the samples to which field samples would be compared for authenticity. We were able to obtain reference samples for most of the varieties collected, but of course could not test samples for which no reference sample existed.

Genetic testing was conducted by the LGC Genomics Laboratory in the United Kingdom. For baseline and CLE2 seed samples, DNA was extracted directly from each seed. At endline, seeds were selected for testing in the same manner as the first two rounds, but each seed was first grown out and leaf clippings taken. Fifty seeds were grown from each sample, and of those that germinated, 20 leaf clippings were taken. They were shipped to LGC Genomics in specialized leaf clipping kits. This process was conducted in Kampala by a seed expert from the National Crops Resources Research Institute (NaCRRI). DNA was then extracted from the leaves.[17] Once all DNA

[16] Initially, a laboratory in the United States was identified to conduct the genetic testing as they had the lowest price per sample for testing. However, many issues were encountered with this laboratory, necessitating a change in laboratories. The LGC Genomics laboratory was identified as being of high quality, but also had a higher cost per sample.

[17] The leaf clipping process was followed at endline because of the lower cost of extracting the DNA. LGC Genomics explained that the two processes would not produce differences in results when analyzing as it was only the DNA extraction method that differed.

was extracted from samples in all rounds, the measurement of the contents of the DNA of each seed was conducted by LGC. Testing was also conducted on the seeds in the reference samples.

For the varieties in Uganda, another research team that included the expert at NaCRRI had already conducted some genetic testing. From this work, they had identified 142 genetic markers on the DNA to which the field samples should be compared. The markers represent different characteristics of the variety (such as color, number of cobs per stalk, etc.).

Once the DNA data was compiled, the raw data was sent to NaCRRI.[18] First, a plant geneticist gathered the information on alleles across the 142 genetic markers and calculated the genetic distance between each of the seeds in all of the *reference samples*. This distance indicated the purity of the reference samples, as well as provided the number to which the comparison to the field samples should be made. Then, within each reference sample, the two seeds with smallest distance between them were chosen as the comparators for each seed in the field samples. The two reference sample seeds with the smallest distance between them were chosen because they were most likely to be an accurate representation of the genetic makeup of the reference sample and because they were least likely to have been affected by issues along the supply chain given that not all reference samples were obtained from breeders. For example, during replication to produce enough seeds to sell on markets, cross-pollination can occur.

Next, the genetic similarity was calculated between each field sample seed and each of the two reference sample seeds from each reference sample. Using these data, for each field sample seed, each reference sample variety was sorted in order of the closest to the furthest genetic distance. At endline, there were ten reference varieties (so 20 reference seeds). So, each sample seed had listed: the variety name of the reference variety seed with the smallest genetic distance measure as well as the distance measure itself, the reference variety seed with the second smallest distance measure as well as the distance measure itself, etc. up to 20 seeds. We also know which maize variety the field sample claims to be from what is stated on the package (or recorded in the tracking data). Using these data, we calculate our outcome measure for the quality of hybrid maize seed at the seed level as the average distance between the field sample seed and the two reference sample seeds that the field sample claims to be. This is the outcome variable in our regression; the smaller the distance the higher the likelihood that the field sample is the variety it claims to be. In our regression, we control for the distance between the two reference sample seeds to capture the 'quality' of information that the reference sample seeds provide. This outcome is a measure of overall quality of the field samples and captures both dimensions of counterfeiting and adulteration. To separate the two, we would need to create arbitrary cutoffs (such as, if x% of seeds in a field sample do not match the reference sample – distance is not more than y% - that the field sample claims to be, it's counterfeit). Cutoffs such as these do not exist even in the genetic testing literature and we are not able to determine what they should be without substantial expertise in

[18] See Appendix C for full details on hybrid maize testing and analysis methods.

this matter. Further, by employing a continuous measure, we are able to use more information and obtain more variation than a simple cutoff.

Table 8.1a displays summary statistics on the average genetic distance between the field sample seeds and the reference sample seeds that the field sample claims to be, by market hub, at baseline, and Table 8.1b does so for endline. 382 samples were collected at baseline and 123 at endline. At baseline, some of the samples went missing between the transfer of the samples from the US laboratory to the UK laboratory, and a handful of samples were not tested because during storage they were severely degraded (for example, some samples had insects in the bags). At endline, all samples collected were tested. We see that overall, the average genetic distance is 0.352 at baseline and 0.371 at endline. These numbers are not statistically different from one another, so we should not conclude that average quality decreased in any way between baseline and endline. These figures are generally considered to be a moderate distance; distances under 0.1 would be considered quite close, and distances above 0.5 are generally considered quite large. There is not much variation across market hubs, and no particular hubs stand out with unusually high or low quality. Some market hubs do have few samples that were collected, however, with only 3 collected in Hoima at baseline and 4 collected in Gulu, 5 in Kasese, and 8 in Iganga at endline. Part of this could be that it was very difficult to predict exactly when products would be available in markets and thus our timing was sometimes off and there were few products to purchase.

Table 8.1a: Average Genetic Distance by Market Hub, Baseline

Market Hub	Average Genetic Distance	Number of Samples Collected	Number of Samples Tested	Number of Seeds Tested
Iganga	0.347	82	73	1,460
Mbale	0.328	43	42	840
Kasese	0.334	10	10	200
Masaka	0.366	84	81	1,620
Hoima	0.356	3	3	60
Masindi	0.365	14	13	260
Mubende	0.362	37	37	740
Kiboga	0.369	16	16	320
Luwero	0.351	35	34	680
Mityana	0.344	24	24	480
Gulu	0.327	34	12	240
Total	**0.352**	**382**	**345**	**6,900**

Table 8.1b: Average Genetic Distance by Market Hub, Endline

Market Hub	Average Genetic Distance	Number of Samples Collected and Tested (all tested)	Number of Seeds Tested
Iganga	0.397	8	160
Mbale	0.327	18	363
Kasese	0.365	5	100
Masaka	0.349	14	282
Hoima	0.340	10	201
Masindi	0.413	19	385
Mubende	0.386	12	260
Kiboga	0.408	8	160
Luwero	0.376	16	325
Mityana	0.378	9	180
Gulu	0.328	4	85
Total	**0.371**	**123**	**2,501**

Table 8.2 below displays results on the impact of the e-verification intervention on the quality of hybrid maize seed. The outcome variable is a decreasing measure of seed quality in that it represents a continuous measure of the average dissimilarity between the seeds in the samples collected in the market from the seeds in the reference sample for that variety. The observations in these models are individual seeds in the 123 field samples, which had close to 20 seeds per sample submitted for analysis.

The table can be read and interpreted as follows. The columns represent the outcome variables (genetic distance). The first row "discount" reports the impact estimate from the treatment – the treatment being the offer of a discount voucher in order to induce more people in treated market locations to purchase EV products compared to market locations in which no community was offered a discount. The coefficient can be interpreted as the percentage point difference in the outcome variable as a result of being offered the discount. The standard error (the precision of the coefficient) is reported in parentheses below. The coefficient and standard error of the control variable of the genetic distance between the two reference sample seeds of the variety that the field sample claims to be are reported in the next two rows. The number of observations in the regression, the R^2 (the proportion of variation in the outcome variable explained by the variables included in the regression), and the average value of the outcome variable in the control group are reported in the last rows.

In column (1), we see that the coefficient on the discount market location dummy variable is negative. It is not statistically significant, but this component of the study was underpowered to detect effects on counterfeiting in several ways. First, we were only able to collect 123 field samples of hybrid maize at endline. Also, the effect of e-verification on average input quality would be derived from the experimentally determined spike in demand for e-verified products that came from providing discount vouchers to farmers in one community near that market. This was a plausible strategy to identify the impact of e-verification on counterfeiting, but it was understood at the design phase that the expected effect for a demand shock of this size would be small and that getting precise estimates of this effect from 120 markets would be challenging. Nonetheless, the result is suggestive of a reduction in counterfeiting. Genetic distance is reduced by 0.007 units of genetic distance as a result of a community served by the market location having been assigned to receive discounts. The mean genetic distance in the market locations that did not have discount communities is 0.376. Thus, e-verification is associated with a reduction in the genetic distance between measured and stated varieties by almost 2 percent. Given that relatively few people redeemed the discount vouchers for hybrid maize seed, the magnitude of the result is not too surprising and provides suggestive evidence that a product assurance scheme can improve the quality of hybrid maize available in the market.

Next, we estimated separately the impact of the discounts to promote take-up of hybrid maize seed separately for E-tag hybrid maize (column 2) and for non-E-tag hybrid maize (column 3). Neither estimate is statistically significant. Moreover, the estimate for a relationship between access to the discount voucher and hybrid seed quality in the tagged sample is very close to zero. This suggests that quality among tagged inputs did not change. In the non-tagged sample, receiving a discount voucher for hybrid maize seed sold in these markets is associated with a reduction in average distance between the field sample seeds and the reference seeds of 0.016, which is a 4.3 percent improvement in this measure of hybrid seed quality. Although this effect is not statistically significant, it documents an association in the data that is consistent with the objectives of the e-verification program – to reduce the prevalence of low-quality inputs outside the E-tag scheme by increasing demand for E-tag products and making lower quality products outside the scheme less profitable. We cannot conclude that e-verification reduced counterfeiting of hybrid seeds outside the program, but these associations document a pattern of quality that fits the market behavior that motivated the e-verification scheme.

Table 8.2: Impact of Discount for E-verified Hybrid Maize Seed on Quality in the Market

	(1) Average Roger's Distance to reference sample (full sample)	(2) Average Roger's Distance to reference sample (tagged sample)	(3) Average Roger's Distance to reference sample (non-tagged sample)
Discount	-0.007	0.001	-0.016
	(0.012)	(0.010)	(0.017)
Distance between two reference samples	0.223	-0.628	0.490***
	(0.146)	(0.447)	(0.141)
Observations	2234	1093	1141
R^2	0.019	0.064	0.110
Control group mean	0.376	0.346	0.398

Notes: Standard errors are clustered at the Market Location level. Distance between two reference samples is the Roger's Distance between the two reference sample seeds that had the smallest distance between them. Regressions are run at the seed level (within a sampled bag of hybrid maize). Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

8.3 Herbicide Quality

Next, we turn to glyphosate herbicide. We first outline the method used to collect the herbicide samples and the number of samples collected, then discuss the way in which samples were tested and the number of samples tested, and finally, the way in which the data was aggregated and analyzed.

As with hybrid maize, we collected samples three times, and at the same time as the hybrid maize samples were collected (August 2014, March 2015, and March 2017). The goal again was to obtain a representative sample of the glyphosate herbicide available in the markets visited by the farmers in our sample, and samples were purchased from the same list of shops. Enumerators thus provided the same information to shopkeepers regarding why they were purchasing samples.

For the first two rounds of data collection, we used the same randomly ordered lists of shops as for hybrid maize seed sample collection. For herbicide, we identified the top 10 brands in terms of market share in each market hub through the same shop surveys. The enumeration team aimed to purchase 8 samples of glyphosate herbicide from each market location, and to purchase brands in a shop that were highest on the list. The instructions were to purchase four samples each from the first two shops, and if one or both of these shops had fewer than four brands, enumerators were to go to the next shop on the list and continue in this manner until 8 samples had been collected or there were not shops left in the market. No brand was to be purchased more than once from the same shop.

Once shops with samples and brands were identified, enumerators first identified the package size to purchase. The two types were: a sealed bottle, and a jerry can, which refers to a large can from which a shop owner would pour smaller quantities of herbicide, generally into a water bottle, if a farmer wanted a smaller quantity than was sold in sealed bottles. If a shop carried more than one size, enumerators were asked to purchase 1 liter bottles, and if necessary, half liter bottles. Once the package type and size were selected, a random number table was used to select the particular bottle to purchase, based on the number of bottles of that brand available in the shop. For each sample purchased, enumerators again recorded, on paper, the shop and brand, the date the bottle was opened/purchased, the date of expiry, and the bottle size and price. Samples purchased were again labeled with sample ID stickers pre-prepared by the research team and a duplicate sticker was included on the paper form. These data were later entered into a CAPI program for analysis. Once the samples were brought back to Kampala, the bottles were thoroughly shaken to mix the contents smoothly, and 100 ml samples of each bottle were measured out into smaller containers for testing.

At baseline, the Government Analytic Laboratory (GAL) performed the testing. The testing method used a High-Pressure Liquid Chromatography (HPLC) machine to analyze the samples (see Appendix D for glyphosate herbicide detailed testing procedures). Some parts of the GAL machine were not working so the study provided funding to obtain these parts. Prior to testing the full sample, samples of known dilution were provided to the lab to test the accuracy of their methods. In addition, duplicate samples were provided to obtain measures of test-retest validity of their approach to measuring quality. These pre-tests were repeated three times at both baseline and endline. The team's project manager frequently visited the laboratory to ensure that tracking of samples was also conducted accurately. Unfortunately, at endline, the pre-tests showed that the testing was of low quality, with substantial inconsistency in estimates from duplicate samples. In addition, the GAL machine had broken by endline. For all of these reasons, the study team decided not to conduct the herbicide testing in Uganda. We sought to export the samples to the United States for testing but were refused an export permit for the samples. The government insisted that the testing of Ugandan products be conducted in country. This delayed the testing of the herbicide samples by a year and a half for the endline samples (and almost three years for the midline samples). The study team was able to identify a chemist who was able to conduct the testing, however, these delays severely affected the quality of the testing data obtained for the endline herbicide testing. In particular, we believe that partial evaporation of water in the herbicide samples led to overconcentration of glyphosate in the samples.

For the endline round of sample collection, only three samples of herbicide were targeted to be purchased from each market location. One sample was to be Kakasa herbicide, and the other two were to be non-tagged herbicide. Shops from the updated shop listing exercise were again ordered randomly within market locations and were to be visited in the specified order. To select a variety, the enumerator was to count the number of different varieties (within the tagged and non-tagged categories) and use a random number table to select a brand. Shops were to be visited in order until one sample of tagged herbicide and two samples of non-tagged herbicide were purchased. Once a brand was selected, a random number table was again used to select the package type from the same two types listed above, attempt to first purchase a one liter bottle and if not possible a half liter bottle, and the specific bottle. Data on the samples collected were recorded in the same manner once again.

For herbicide, a 'reference' was also needed to which field samples would be compared. The reference in this case was a "glyphosate standard", which has a known concentration of glyphosate content. The study team purchased these standards for the laboratories. For both rounds of testing, the 100 ml herbicide sample was diluted in water before testing. See Appendix D for full details on testing methods for glyphosate herbicide.

Once the testing was complete, the laboratories provided the study team with the measured concentration of glyphosate in each sampled bottle. We then compared this measured concentration with the glyphosate concentration that was stated on the bottle. Our outcome measure is the measured concentration divided by the stated concentration. At baseline, two tests were performed on the same sample and an average of the two measured concentrations is used before dividing it by the stated concentration, and at endline, only one test was conducted per sample. As with hybrid maize seeds, this method avoids selecting arbitrary cutoffs and makes use of all the variation and information in the data, and thus represents quality from several dimensions including counterfeiting, adulteration, and storage issues along the supply chain.

Table 8.3a displays summary statistics on the ratio of measured to stated glyphosate concentration, by market hub, at baseline, and Table 8.3b does so for endline. We were able to test almost all herbicide samples at baseline and endline. A few samples were not tested because they may have been misplaced, mis-labeled, or degraded. We see that overall, the average ratio of measured to stated glyphosate concentration is 0.896 at baseline and 1.288 at endline. First, we should not necessarily conclude that the prevalence of poor-quality herbicide is low. We also should not conclude that herbicide quality increased from baseline to endline. Both the baseline and endline averages are high but reflect some degree of time that elapsed between the collection and testing of the samples. During this time, other parts of the herbicide solution evaporate and leave a higher concentration of glyphosate in the bottle. This is quite evident in the endline samples as all of the ratios are higher than one, indicating a very high degree of evaporation.[19] There is not much variation across market hubs, and while Mubende looks different as it has a ratio higher than one, this may reflect the time between collection and testing or storage conditions. Some market hubs do have few samples that were collected, however, with only 4 collected in Mbale at baseline. This is again because it was very difficult to predict exactly when products would be available in markets and thus our timing was sometimes off and there were few products to purchase.

Table 8.3a: Quality of Herbicide Samples, Baseline

Market Hub	Average Measured/Stated Concentration	Number of Samples Collected	Number of Samples Tested
Iganga	0.819	37	33
Mbale	0.930	4	3
Kasese	0.929	31	27
Masaka	0.988	102	94
Hoima	0.722	44	42
Masindi	0.725	61	60
Mubende	1.172	88	81
Kiboga	0.852	37	36
Luwero	0.798	68	63
Mityana	0.852	71	62
Gulu	0.782	13	13
Total	**0.896**	**556**	514

[19] The samples were also not stored as recommended. Herbicide should be stored at 4 degrees centigrade in order not to evaporate and should always be properly sealed. Not all of the samples that were purchased were originally sealed, and there was nowhere that so many samples could be stored at that temperature.

Table 8.3b: Quality of Herbicide Samples, Endline

Market Hub	Average Measured/Stated Concentration	Number of Samples Collected	Number of Samples Tested
Iganga	1.495	27	21
Mbale	1.478	23	13
Kasese	1.440	27	23
Masaka	0.998	45	43
Hoima	1.173	39	30
Masindi	1.390	36	27
Mubende	1.300	42	39
Kiboga	1.030	27	25
Luwero	1.462	36	30
Mityana	1.473	36	35
Gulu	1.131	23	18
Total	**1.288**	**358**	**304**

Table 8.4 presents estimates of the impact of the e-verification intervention on herbicide quality. We estimate the same regression as in equation (4) but of course without the Roger's distance between the reference seed samples as a control. The table can be read in the same way as that for hybrid maize seeds, but with the outcome variable in the columns being the measured glyphosate content divided by the stated glyphosate content. The regression coefficient for the full sample (column 1) is negative, meaning that the ratio of measured to stated concentration is lower in discount community market locations. However, this point estimate is not statistically significant. Column 2 presents the estimates for the impact of the increased demand for e-verified glyphosate on the quality of herbicide samples that bore the E-tag label. There is no significant effect, but the estimate is positive. For herbicide samples that were not in the E-tag program, the estimated effect of the discount on herbicide quality is negative but is not statistically significant. Given the degraded conditions of the herbicide samples, while we present results here to be transparent, we do not believe that any conclusions can be drawn from these data.

Table 8.4: Impact of Discount on the Quality of Herbicide

	(1) Ratio of actual-to-stated glyphosate content (full sample)	(2) Ratio of actual-to-stated glyphosate content (tagged sample)	(3) Ratio of actual-to-stated glyphosate content (non-tagged sample)
Discount treatment assignment	-0.028	0.086	-0.044
	(0.061)	(0.125)	(0.066)
Observations	286	43	241
R^2	0.001	0.009	0.002
Control group mean	1.312	1.190	1.329

Notes: Standard errors are clustered at the Market Location level. Regressions are run at the sample level for glyphosate herbicide. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

9. Conclusion and Recommendations

The USAID FTF Ag Inputs activity, in collaboration with the Government of Uganda and other partners, supported implementation of the Kakasa E-tag and Ag-verify e-verification initiatives to address widespread concern that agricultural input markets were plagued with counterfeit inputs, leading to depressed demand for high quality inputs and impeding agricultural productivity and the profits of smallholders.

The impact evaluation used a randomized encouragement design to experimentally induce higher take-up of the e-verification program among farmers in randomly selected encouragement villages. The evaluation results from the reduced form models show that the encouragement program itself was effective at promoting increases in the use of hybrid maize seed, glyphosate herbicide, any herbicide and inorganic fertilizer of 2.5-4.3 percent. These modest effects demonstrate the validity of the encouragement intervention; the introduction of the encouragement campaign improved many of the primary outcomes related to input use in the study.

Of greater importance and policy relevance, however, is the effectiveness of the e-verification initiative. The IV models exploit the randomized encouragement intervention to estimate the causal effect of e-verification on input use and farm productivity and profits. In these estimates, e-verification caused a statistically significant increase in the use of hybrid maize seed by 5.4 percentage points in first season 2017. This effect is large given that only 10.5 percent of households used hybrid maize seed in the control group at endline. The impact on use of any maize seed is an increase by 5.6 percentage points. E-verification caused an increase in take-up of glyphosate herbicide of 7.7 percentage points (compared to a 43.8 percent adoption rate in the control group) and of any herbicide by 6.6 percentage points. Use of inorganic fertilizer increased 3.2 percentage points, but this effect was only weakly significant. Use of any fertilizer increased 5.3 percent. The pattern of impacts also indicates that the new adoption of hybrid maize varieties came through cultivation of hybrid maize by households that would not otherwise have grown maize, so the effect is one of hybrid maize expansion rather than substitution of hybrid maize for conventional maize, in area planted.

It is important to note that the effect of e-verification on fertilizer is a spillover effect, since plans to include labeled small-pack fertilizer in the e-verification scheme were not implemented during the study period. Instead, these results indicate the impact of having purchased e-verified hybrid maize seed or glyphosate herbicide on use of fertilizer. This positive effect on fertilizer use may operate through several channels: (i) farmers shopping for e-verified maize seed and herbicide decide to also purchase fertilizer while in the input shops; (ii) the e-verification scheme increased confidence among farmers that counterfeiting of all inputs must be declining; or (iii) purchases of hybrid maize seed led to higher fertilizer demand because the two are complementary inputs.

In second season 2016, the encouragement intervention led to smaller increases in purchases of e-verified products, likely because fewer e-verified products were on the shelf in that first season of the program. Nonetheless, e-verification lead to large and significant increase in the use of hybrid maize seed of 32.2 percent in second season 2016.[20]

We find no evidence that the increases in input use caused by e-verification led to increases in agricultural productivity or profits. There is a significant effect of the program on the number of crops grown in second season 2016, but not other effects on farm output. It is possible that these effects could take more time to be realized. It is also possible that the inputs taken up are not as productive and profitable on average as expected. Suri (2011) showed that there is substantial heterogeneity in the profitability of agricultural inputs.

The e-verification scheme led to an increase in beliefs about whether the quality of conventional maize seed or hybrid maize seed is sometimes lowered in the market, which may reflect the increase in awareness about the presence of the e-verification authentication scheme. Also, impacts of e-verification on the input use appears to be driven mostly by men in the sample. Impacts on use of hybrid maize seed were somewhat large for the poorest (based on asset quintile) households in the sample, but this adoption effect for the poor did not lead to positive impacts on yield or net income for poor households. Finally, there were no statistically significant impacts of e-verification on the quality of maize seed or herbicide in the market (a measure of counterfeiting), though associations with maize seed quality suggests that it is plausible that the scheme could eventually lead to changes in the input market that would drive out low quality producers.

We have some recommendations based on these results and other lessons learned from four years of work on the evaluation. First, the e-verification scheme should be expanded to promote further increases in adoption of high-quality inputs. In addition, E-tag and Ag-verify are currently designed to provide product assurance (what you see on the label is what you get), but not quality assurance, as there is no testing or enforcement mechanism to assure that the products are actually of high quality. E-verification should be expanded to include such testing and enforcement components to further increase confidence in input markets and foster greater profitability.

[20] It is surprising that the impact of e-verification on use of any maize seed is a 22.2 percent increase in second season 2016, since 93.6 percent of farmers in the control group grew maize that season. We are reviewing the estimates to check if this is a valid result.

References

Akerlof, George. 1970. “The market for lemons.” *Quarterly Journal of Economics* 84 (3): 488–500.

Anderson, Michael L. 2008. “Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of the American Statistical Association* 103 (484): 1481–1495.

Ashour, Maha, Daniel O. Gilligan, Jessica B. Hoel, and Naureen Karachiwalla. 2018. “Do Beliefs about Herbicide Quality Correspond with Actual Quality in Local Markets?” Accepted for publication at *Journal of Development Studies*.

Ashour, Maha, Lucy Billings, Daniel O. Gilligan and Naureen Karachiwalla. 2015. *An Evaluation of the Impact of E-verification on Counterfeit Agricultural Inputs and Technology Adoption in Uganda: Baseline Report.* Washington DC: International Food Policy Research Institute.

Bhaskara, B. L. and Nagaraja, P. (2006). Direct sensitive spectrophotometric determination of Glyphosate by using Ninhydrin as a chromogenic reagent in formulation and environmental water samples. Helv. Chim. Acta 89: 2686-2693.

Bloch, Peter, Bruce Kisitu, and Andrew Gita. 2013. Report on a Program Pilot to Test Counter-Measures Designed to Mitigate the Sales and Use of Counterfeit Crop Protection Products in Sub-Saharan Africa. IFDC.

Bold, T., Kaizzi, K.C., Svensson, J. and Yanagizawa-Drott, D., 2017. Lemon technologies and adoption: measurement, theory and evidence from agricultural markets in Uganda. *The Quarterly Journal of Economics*, *132*(3), pp.1055-1100.

Bruhn, Miriam, and David McKenzie. 2009. “In pursuit of balance: Randomization in practice in development field experiments.” *American Economic Journal: Applied Economics* 1 (4): 200–232.

Cardenas, Juan Camilo, and Jeffrey Carpenter. 2013. “Risk attitudes and economic well-being in Latin America.” *Journal of Development Economics* 103: 52–61.

Duflo, Esther, and Emmanuel Saez. 2003. “The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment.” *Quarterly Journal of Economics* 118 (3): 815–842.

Duflo, Ester, Michael Kremer, and Jonathan Robinson. 2011. “Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya.” *American Economic Review* 101 (6): 2350–2390.

Gilligan, Daniel O., Neha Kumar, Scott McNiven, J. V. Meenakshi, and Agnes Quisumbing. 2013. *Bargaining Power and Biofortification: The Role of Gender in Adoption of Orange Sweet Potato in Uganda*. Washington, DC: International Food Policy Research Institute.

Goldman, Megan. 2008. http://www.stat.berkeley.edu/~mgoldman/Section0402.pdf (downloaded November 29, 2014).

Haile, Beliyou A. 2011. “Promoting HIV testing and safe sexual behavior: evidence from a field experiment in Ethiopia.” Job market paper, Department of Economics, Columbia University, New York.

Holt, Charles A., and Susan K. Laury. 2002. “Risk aversion and incentive effects.” *American Economic Review* 92 (5): 1644–1655.

Jamison, Julian, Dean Karlan, and Jonathan Zinman. 2012. “Measuring risk and time preferences and their connections with behavior.” Mimeo.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman. 2001. “Moving to opportunity in Boston: Early results of a randomized mobility experiment.” *Quarterly Journal of Economics* 116: 607–654.

Monitor Deloitte. 2014. “Counterfeiting in African Agriculture Inputs—Challenges & Solutions.” Draft PowerPoint presentation, Bill & Melinda Gates Foundation, Nairobi.

Nagaraja, P., Shrestha, A. K., Shivakumar, A., & Al-Tayar, N. G. S. (2011). Molybdate assisted ninhydrin based sensitive analytical system for the estimation of drugs containing amine group. *J Food Drug Anal*, 19(1), 85-93.

Semagn, K., Beyene, Y., Makumbi, D., Mugo, S., Prasanna, B.M., Magorokosho, C. and Atlin, G. (2012). Quality control genotyping for assessment of genetic identity and purity in diverse tropical maize inbred lines. *Theoretical and Applied Genetics* 125: 1487-501.

Suri, Tavneet. 2011. “Selection and comparative advantage in technology adoption.” *Econometrica* 79 (1): 159–209.

Svensson, Jakob, David Yanagizawa-Drott, and Tessa Bold. 2013. “The market for (fake) agricultural inputs: Short summary of research project.” Mimeo.

Thornton, Rebecca L. 2008. “The demand for, and impact of, learning HIV status.” *American Economic Review* 98 (5): 1829–1863.

Appendix A: Robustness Check – Specifications Including Baseline Household Controls

Table A1: Reduced-form Impacts on the Use of Input in First Season 2017, With baseline HH controls

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Encouragement	0.032***	0.030***	0.044***	0.037***	0.017*	0.030**
	(0.010)	(0.011)	(0.013)	(0.013)	(0.010)	(0.012)
Number of observations	2,026	2,024	1,954	1,994	1,994	1,981
R-squared	0.178	0.123	0.445	0.450	0.187	0.267
Control group mean	0.105	0.888	0.438	0.454	0.102	0.150

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the average baseline outcome variables for first season 2014 and 2015, and for second season 2013 and 2015. Baseline HH control variables included are: the age, gender (male) and literacy of both the household head and the primary agricultural decision maker, household size, total land area, and an index of household assets (which includes durable assets and productive assets constructed using principal components analysis). When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table A2: Reduced-form Impacts on the Use of Input in Second Season 2016, With baseline HH controls

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
Encouragement	0.018**	0.019	0.005	0.009	0.002	0.001
	(0.009)	(0.012)	(0.014)	(0.013)	(0.008)	(0.010)
Number of observations	1,942	1,959	1,868	1,930	1,930	1,930
R-squared	0.188	0.153	0.420	0.426	0.143	0.170
Control group mean	0.076	0.808	0.350	0.372	0.062	0.100

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the average baseline outcome variables for first season 2014 and 2015, and for second season 2013 and 2015. Baseline HH control variables included are: the age, gender (male) and literacy of both the household head and the primary agricultural decision maker, household size, total land area, and an index of household assets (which includes durable assets and productive assets constructed using principal components analysis). When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table A3: Reduced-form Impacts on Farm Production Outcomes in First Season 2017, With baseline HH controls

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income from farm work (million UGX)	(5) Log maize yield in representative maize plots (tons per acre)	(6) Log value of farm production in representative maize plots (million UGX per acre)	(7) IHS* of net income from farm work in representative maize plots (million UGX)
Encouragement	0.033	-0.026	-0.009	0.139	-0.035	-0.006	-0.401
	(0.045)	(0.042)	(0.037)	(0.245)	(0.041)	(0.075)	(0.246)
Number of observations	2,039	1,641	1,932	1,962	1,611	1,638	1,616
R-squared	0.274	0.304	0.222	0.154	0.273	0.338	0.168
Control group mean	4.199	4.905	12.497	12.246	5.891	13.411	11.895

Notes: IHS* is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the average baseline outcome variables for first season 2014 and 2015, and for second season 2013 and 2015. Baseline HH control variables included are: the age, gender (male) and literacy of both the household head and the primary agricultural decision maker, household size, total land area, and an index of household assets (which includes durable assets and productive assets constructed using principal components analysis). When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table A4: Reduced-form Impacts on Farm Production Outcomes in Second Season 2016, With baseline HH controls

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income from farm work (million UGX)
Encouragement	0.108**	0.016	0.004	0.299
	(0.047)	(0.052)	(0.031)	(0.185)
Number of observations	2,034	1,498	1,864	1,888
R-squared	0.283	0.309	0.220	0.145
Control group mean	3.629	4.906	12.671	12.901

Notes: IHS* is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the average baseline outcome variables for first season 2014 and 2015, and for second season 2013 and 2015. Baseline HH control variables included are: the age, gender (male) and literacy of both the household head and the primary agricultural decision maker, household size, total land area, and an index of household assets (which includes durable assets and productive assets constructed using principal components analysis). When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table A5: Impacts on the Use of Agricultural Inputs in First Season 2017 with IV Specifications, With Baseline HH Controls

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
First stage						
Encouragement	0.566***	0.563***	0.560***	0.560***	0.561***	0.562***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
First stage F-statistic	1,140.7	1,132.1	1,056.5	1,088.8	1,102.9	1,098.1
Second stage						
Heard of Etag - predicted	0.057***	0.054***	0.078***	0.068***	0.030*	0.053***
	(0.017)	(0.018)	(0.023)	(0.022)	(0.016)	(0.020)
Number of observations	2,024	2,023	1,954	1,993	1,993	1,981
R-squared	0.057	0.022	0.163	0.171	0.086	0.084
Control group mean	0.105	0.888	0.438	0.454	0.102	0.150

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the average baseline outcome variables for first season 2014 and 2015, and for second season 2013 and 2015. Baseline HH control variables included are: the age, gender (male) and literacy of both the household head and the primary agricultural decision maker, household size, total land area, and an index of household assets (which includes durable assets and productive assets constructed using principal components analysis). When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table A6: Impacts on the Use of Agricultural Inputs in Second Season 2016 with IV Specifications, With Baseline HH Controls

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
First stage						
Encouragement	0.563***	0.562***	0.567***	0.561***	0.562***	0.562***
	(0.017)	(0.017)	(0.018)	(0.017)	(0.017)	(0.017)
First stage F-statistic	1,084.8	1,078.7	1,041.1	1,048.2	1,061.3	1,058.0
Second stage						
Heard of Etag - predicted	0.032**	0.033*	0.009	0.016	0.003	0.002
	(0.015)	(0.020)	(0.023)	(0.023)	(0.014)	(0.017)
Number of observations	1,942	1,959	1,868	1,930	1,930	1,930
R-squared	0.051	0.014	0.135	0.136	0.062	0.045
Control group mean	0.076	0.808	0.350	0.372	0.062	0.100

Notes: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the average baseline outcome variables for first season 2014 and 2015, and for second season 2013 and 2015. Baseline HH control variables included are: the age, gender (male) and literacy of both the household head and the primary agricultural decision maker, household size, total land area, and an index of household assets (which includes durable assets and productive assets constructed using principal components analysis). When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table A7: Impacts on Farm Production Outcomes in First Season 2017 with IV Specifications, With Baseline HH Controls

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income from farm work (million UGX)	(5) Log maize yield in representative maize plots (tons per acre)	(6) Log value of farm production in representative maize plots (million UGX per acre)	(7) IHS* of net income from farm work in representative maize plots (million UGX)
First stage							
Encouragement	0.564***	0.590***	0.569***	0.564***	0.595***	0.598***	0.596***
	(0.017)	(0.019)	(0.017)	(0.017)	(0.019)	(0.018)	(0.018)
First stage F-statistic	1,128.7	986.1	1,116.6	1,087.8	1,024.4	1,069.3	1,056.8
Second stage							
Heard of Etag - predicted	0.056	-0.044	-0.016	0.247	-0.058	-0.010	-0.673*
	(0.077)	(0.068)	(0.063)	(0.417)	(0.066)	(0.121)	(0.395)
Number of observations	2,035	1,641	1,932	1,962	1,611	1,638	1,616
R-squared	0.117	0.065	0.025	0.018	0.030	0.023	0.015
Control group mean	4.199	4.905	12.497	12.246	5.891	13.411	11.895

Notes: IHS* is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the average baseline outcome variables for first season 2014 and 2015, and for second season 2013 and 2015. Baseline HH control variables included are: the age, gender (male) and literacy of both the household head and the primary agricultural decision maker, household size, total land area, and an index of household assets (which includes durable assets and productive assets constructed using principal components analysis). When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * $p < 0.1$, ** $p < .05$, *** $p < 0.01$.

Table A8: Impacts on Farm Production Outcomes in Second Season 2016 with IV Specifications, With Baseline HH Controls

	(1) Number of crops grown	(2) Log maize yield (tons per acre)	(3) Log value of farm production (million UGX per acre)	(4) IHS* of net income from farm work (million UGX)
First stage				
Encouragement	0.563***	0.559***	0.563***	0.563***
	(0.017)	(0.019)	(0.018)	(0.018)
First stage F-statistic	1,127.8	876.1	1,006.7	1,017.6
Second stage				
Heard of Etag - predicted	0.184**	0.028	0.008	0.531*
	(0.080)	(0.089)	(0.053)	(0.317)
Number of observations	2,030	1,498	1,864	1,888
R-squared	0.106	0.066	0.023	0.029
Control group mean	3.629	4.906	12.671	12.901

Notes: IHS* is inverse hyperbolic sine transformation. Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the average baseline outcome variables for first season 2014 and 2015, and for second season 2013 and 2015. Baseline HH control variables included are: the age, gender (male) and literacy of both the household head and the primary agricultural decision maker, household size, total land area, and an index of household assets (which includes durable assets and productive assets constructed using principal components analysis). When a household is missing a baseline household value, the missing value is replaced with the mean value for the sample, and a dummy variable is included for all households for which the baseline value is missing. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Appendix B: Robustness Check – IV Specifications with Purchase of any E-tag/Ag-verify Input as Endogenous variable

Table B1: Impact of E-verification on Use of Agricultural Inputs in First Season 2017, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
First stage						
Encouragement	0.200***	0.200***	0.186***	0.187***	0.189***	0.189***
	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)
First stage F-statistic	102.7	102.3	97.4	100.9	104.0	104.2
Second stage						
Predicted E-tag /Ag-verify purchase	0.159***	-0.001	0.300***	0.291***	0.147***	0.229***
	(0.052)	(0.020)	(0.071)	(0.069)	(0.050)	(0.063)
Number of observations	1,799	1,799	1,960	1,987	1,987	1,987
R-squared	0.161	0.083	0.441	0.446	0.147	0.221
Mean dependent variable	0.124	0.986	0.461	0.477	0.110	0.166

Note: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table B2: Impact of E-verification on Use of Agricultural Inputs in Second Season 2016, IV Specifications

	(1) Used hybrid maize	(2) Used any maize	(3) Used glyphosate herbicide	(4) Used any herbicide	(5) Used inorganic fertilizer	(6) Used any fertilizer
First stage						
Encouragement	0.080***	0.079***	0.068***	0.070***	0.071***	0.071***
	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)
First stage F-statistic	51.9	51.2	47.6	50.8	51.1	51.5
Second stage						
Predicted E-tag /Ag-verify purchase	0.322**	0.222**	0.224	0.316*	0.165	0.191
	(0.126)	(0.101)	(0.189)	(0.182)	(0.107)	(0.133)
Number of observations	1,646	1,646	1,925	1,987	1,987	1,987
R-squared	0.161	0.095	0.402	0.413	0.096	0.140
Mean dependent variable	0.097	0.936	0.341	0.369	0.061	0.098

Note: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table B3: Impact of E-verification on Farm Production Outcomes in First Season 2017, IV Specifications

	(1) Number of crops grown	(2) Maize yield, tons per acre	(3) Value of farm production, million UGX per acre	(4) Net income from farm work, million UGX
First stage				
Encouragement	0.190***	0.202***	0.191***	0.235***
	(0.013)	(0.014)	(0.013)	(0.017)
First stage F-statistic	106.8	103.1	104.8	100.2
Second stage				
Predicted E-tag /Ag-verify purchase	0.312	-0.605	-0.019	-0.373
	(0.229)	(0.576)	(0.101)	(0.990)
Number of observations	2,033	1,777	1,971	1,314
R-squared	0.231	0.085	0.084	0.175
Mean dependent variable	4.126	0.824	0.235	1.452

Note: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Table B4: Impact of E-verification on Farm Production Outcomes in Second Season 2016, IV Specifications

	(1) Number of crops grown	(2) Maize yield, tons per acre	(3) Value of farm production, million UGX per acre	(4) Net income from farm work, million UGX
First stage				
Encouragement	0.069***	0.082***	0.074***	0.088***
	(0.007)	(0.008)	(0.007)	(0.010)
First stage F-statistic	51.5	49.4	53.1	39.2
Second stage				
Predicted E-tag /Ag-verify purchase	1.563**	0.728	-0.101	0.292
	(0.739)	(0.944)	(0.171)	(1.730)
Number of observations	2,033	1,549	1,918	1,248
R-squared	0.215	0.086	0.151	0.235
Mean dependent variable	3.595	0.761	0.293	1.566

Note: Standard errors are reported in parentheses. Regressions include a dummy variable for whether the household did not own a phone during the first CLE, and controls for the baseline outcome variables for first season 2014 and second season 2013. Regressions also include market location dummy variables, and standard errors are clustered at the village level. Regressions are also weighted for the likelihood of sampling households with and without phones. Stars denote statistical significance as follows: * p < 0.1, ** p < .05, *** p < 0.01.

Appendix C: Technical Appendix on Quality Testing for Hybrid Maize Seed

This technical appendix describes the methodology used for lab tests conducted by LGC Genomics Laboratory in the United Kingdom to determine the genetic composition of hybrid maize seed samples. In this description, reference sample refers to a sample of seeds procured directly from the seed producing company. These seeds are representative of those that are sent to contracted farmers to mass produce the seed for sale in the market. Field sample refers to seeds purchased in a package or from a bin in shops in local markets as part of the surveys. Field samples represent the seed that farmers have available to them in one of their local markets, usually the closest market.

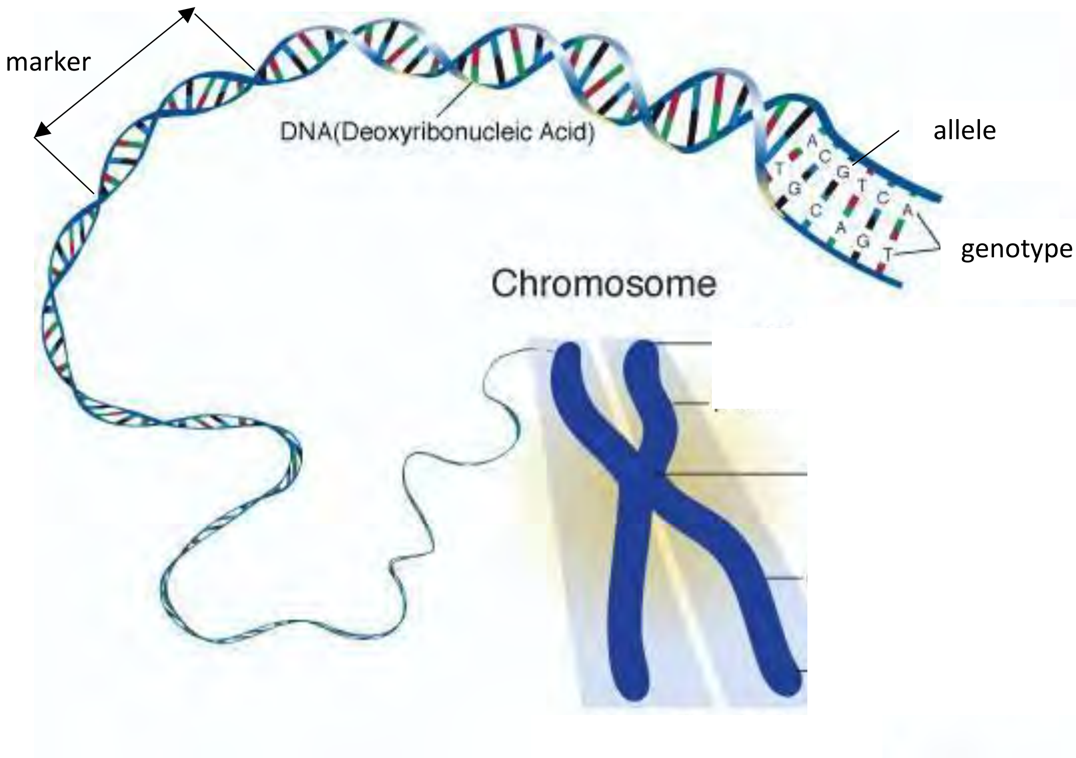
Testing procedures differed slightly at baseline and endline:[21]

- Baseline: Genetic material was extracted directly from seeds for analysis.
    - From each field sample, 20 seeds were randomly selected. To select these seeds, the entire sample was poured onto a table that was divided into four quadrants using tape. The seeds were then mixed across the quadrants and a protocol was followed to attempt a random draw of seeds from each quadrant in order to provide a sample of 20 seeds for lab analysis.

- Endline: Genetic material was extracted from leaf clippings of seeds that had been planted and grown for analysis.
    - From each field sample, 50 seeds were randomly selected. To select these seeds, the entire sample was poured onto a table that was divided into four quadrants using tape. The seeds were then mixed across the quadrants and a protocol was followed to attempt a random draw of seeds from each quadrant in order to provide a sample of 50 seeds for lab analysis.
    - All 50 seeds selected per field sample were planted and grown in separate pots in a greenhouse by the National Crops Resources Research Institute (NaCRRI) in Kampala, Uganda.
    - Once grown, leaves were clipped from a randomly selected set of 20 seeds that germinated out of the 50 seeds planted. Leaves were clipped when they reached leaf stage 3-4 based on a protocol developed by LGC.[22] Leaf samples were sent to LGC in their "Plant Sample Kit".

[21] The two procedures are equivalent in terms of data produced; leaf clippings are easier to transport and are a cheaper method by which to extract DNA. This is why the procedure was changed between baseline and endline.
[22] LGC. Plant Sample Kit. https://biosearch-cdn.azureedge.net/assetsv6/Plant-leaf-kit.pdf

Figure 1: Illustration of DNA and its components



Source: National Human Genome Research Institute. https://www.genome.gov/genetics-glossary/Chromosome.

- LGC extracted the DNA from both baseline and endline samples and used Kompetitive Allele Specific PCR (KASP), where PCR is Polymerase Chain Reaction (a way of making many copies of DNA, which is needed to do the testing). Figure 1 illustrates the genetic components for this discussion.
    - An **allele** is a heritable unit of a gene that determines a particular feature of an organism. There are four possible alleles in DNA – C, T, A, G (four types of proteins). A **genotype** is a combination of two of these four alleles. Populations with many similar alleles have more similar traits. This indicates that they are closely related and have a recent common ancestor.
    - Genetic **marker** – the location of a string of adjacent genotypes on a **chromosome.**

  In this study, 142 markers were used. Markers were selected based on recommendations for routine low-cost genotyping (Semagn et al. 2012) and also have a good distribution across all chromosomes; these were used in previous genetic studies for seeds in Uganda. LGC provided the genotype for each seed and each marker.

- Data was cleaned and imported into Power-Marker software. For each marker, there are only two possible alleles (two alleles make a genotype). A matrix of allele frequencies was created to note the proportion of each of the two alleles that occur (proportion of C, T, A,

G) in each marker for each seed (*C/2*, *T/2*, *A/2*, and *G/2*). This is the allele frequency – the proportion of times an allele occurs in a seed for a particular marker. See Figure 2. For example, for marker PHM10404_8, there are two possible alleles – C and G. For each seed (listed in each column), the matrix will tell us the proportion of Cs and Gs recorded (essentially, whether they are both Cs, both Gs, or one of each).

Figure 2: Allele frequencies

	A	B	C	D	E	F	G	H	I
1		**Seed**	**5501-52**	**5501-55**	**5501-56**	**5501-59**	**5501-60**	**5501-62**	**5501-63**
2	**Marker**	**Allele**							
3	**PHM10404_8**	C	0.5000	0.5000	0.5000	1.0000	1.0000	0.5000	1.0000
4	**PHM10404_8**	G	0.5000	0.5000	0.5000	0.0000	0.0000	0.5000	0.0000
5	**PHM11985_27**	A	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000
6	**PHM11985_27**	G	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
7	**PHM1218_6**	A	0.5000	0.0000	0.0000	0.0000	0.5000	0.0000	0.5000
8	**PHM1218_6**	G	0.5000	1.0000	1.0000	1.0000	0.5000	1.0000	0.5000
9	**PHM12979_9**	A	0.5000	0.5000	0.0000	1.0000	0.5000	1.0000	0.5000
10	**PHM12979_9**	G	0.5000	0.5000	1.0000	0.0000	0.5000	0.0000	0.5000
11	**PHM13673_53**	A	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
12	**PHM13673_53**	G	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
13	**PHM13675_17**	A	0.5000	0.0000	0.0000	0.0000	0.5000	0.5000	0.0000
14	**PHM13675_17**	T	0.5000	1.0000	1.0000	1.0000	0.5000	0.5000	1.0000
15	**PHM13687_14**	A	1.0000	0.0000	0.0000	0.5000	0.0000	0.5000	1.0000
16	**PHM13687_14**	G	0.0000	1.0000	1.0000	0.5000	1.0000	0.5000	0.0000
17	**PHM13696_9**	A	1.0000	0.5000	1.0000	0.5000	1.0000	0.5000	0.5000
18	**PHM13696_9**	G	0.0000	0.5000	0.0000	0.5000	0.0000	0.5000	0.5000

- Then the genetic distance between each seed to every other seed (across field samples and reference samples) is calculated using the Roger's Distance formula:

$$D_R = \frac{1}{L}\sqrt{\sum_{ul}(X_{ul} - Y_{ul})^2/2}$$

where *X* and *Y* represent two different seeds populations for the *L* loci (markers – 142) being studied. Let X_{ul} represent the u^{th} allele frequency (proportion) at the l^{th} locus. Roger's distance provides a measure of the degree of genetic similarity between two seeds. A smaller Roger's distance indicates that two seeds are more similar.

- Calculation of the Roger's distance produces a matrix with the genetic distance (between 0 and 1) between each seed and every other seed (see Figure 3). Note, rows and columns are individual seeds (so it is a symmetric matrix).

Figure 3: Genetic Distance Matrix

	A	CQY	CQZ	CRA	CRB	CRC	CRD	CRE	CRF	CRG	CRH	CRI	CRJ	CRK	CRL	CRM	CRN	CRO	CRP
1	ID	31690-74	31690-76	31690-77	31690-79	31690-81	31690-82	31690-86	31690-89	31690-90	5501-56	5501-85	5503-80	5503-95	5513-77	5513-82	5515-62	5515-65	5516-75
2491	31690-66	0.39513	0.20372	0.41465	0.20851	0.231	0.231	0.19378	0.35564	0.21779	0.52129	0.51939	0.44009	0.44898	0.40379	0.41226	0.33119	0.31119	0.33856
2492	31690-69	0.44455	0.25922	0.43101	0.2132	0.20851	0.19881	0.19881	0.385	0.18329	0.53439	0.52129	0.47673	0.48086	0.40133	0.40986	0.34002	0.27404	0.36388
2493	31690-71	0.385	0.18329	0.44677	0.19881	0.22228	0.2132	0.24752	0.36659	0.20851	0.51748	0.51939	0.4577	0.47465	0.41346	0.41703	0.30639	0.31119	0.33267
2494	31690-74	0	0.36116	0.44455	0.4001	0.4264	0.4264	0.39762	0.43784	0.39513	0.53439	0.53991	0.54265	0.54628	0.45227	0.45985	0.37327	0.38755	0.39513
2495	31690-76	0.36116	0	0.40256	0.20372	0.22668	0.19881	0.22668	0.33563	0.2132	0.52318	0.51748	0.47257	0.47673	0.41106	0.41939	0.31592	0.32057	0.32364
2496	31690-77	0.44455	0.40256	0	0.44677	0.4264	0.44009	0.39762	0.45117	0.43784	0.50394	0.50589	0.49703	0.48901	0.38628	0.41465	0.39886	0.44009	0.41939
2497	31690-79	0.4001	0.20372	0.44677	0	0.20372	0.231	0.231	0.38242	0.18861	0.53624	0.53068	0.46199	0.47047	0.41346	0.41226	0.3371	0.31748	0.32668
2498	31690-81	0.4264	0.22668	0.4264	0.20372	0	0.18861	0.21779	0.37459	0.231	0.53068	0.52129	0.48495	0.49303	0.44344	0.45553	0.32819	0.2881	0.35285
2499	31690-82	0.4264	0.19881	0.44009	0.231	0.18861	0	0.23524	0.36927	0.231	0.55614	0.54718	0.49303	0.50099	0.44344	0.45985	0.34002	0.30151	0.35285
2500	31690-86	0.39762	0.22668	0.39762	0.231	0.21779	0.23524	0	0.36927	0.20372	0.51939	0.50979	0.47673	0.48086	0.39638	0.40501	0.32819	0.2881	0.35285
2501	31690-89	0.43784	0.33563	0.45117	0.38242	0.37459	0.36927	0.36927	0	0.36659	0.53254	0.52694	0.4788	0.48698	0.43671	0.44455	0.39638	0.40986	0.40256
2502	31690-90	0.39513	0.2132	0.43784	0.18861	0.231	0.231	0.20372	0.36659	0	0.51365	0.49603	0.46625	0.47047	0.39886	0.40256	0.32516	0.31119	0.32668
2503	5501-56	0.53439	0.52318	0.50394	0.53624	0.53068	0.55614	0.51939	0.53254	0.51365	0	0.263	0.52881	0.55078	0.46413	0.48189	0.49901	0.53808	0.47361
2504	5501-85	0.53991	0.51748	0.50589	0.53068	0.52129	0.54718	0.50979	0.52694	0.49603	0.263	0	0.53439	0.54537	0.44898	0.47569	0.49703	0.52881	0.45877
2505	5503-80	0.54265	0.47257	0.49703	0.46199	0.48495	0.49303	0.47673	0.4788	0.46625	0.52881	0.53439	0	0.17782	0.46306	0.4577	0.45662	0.48086	0.44455
2506	5503-95	0.54628	0.47673	0.48901	0.47047	0.49303	0.50099	0.48086	0.48698	0.47047	0.55078	0.54537	0.17782	0	0.47152	0.47047	0.46942	0.48901	0.45336
2507	5513-77	0.45227	0.41106	0.38628	0.41346	0.44344	0.44344	0.39638	0.43671	0.39886	0.46413	0.44898	0.46306	0.47152	0	0.20128	0.4333	0.45227	0.39886
2508	5513-82	0.45985	0.41939	0.41465	0.41226	0.45553	0.45985	0.40501	0.44455	0.40256	0.48189	0.47569	0.4577	0.47047	0.20128	0	0.43896	0.45985	0.41226
2509	5515-62	0.37327	0.31592	0.39886	0.3371	0.32819	0.34002	0.32819	0.39638	0.32516	0.49901	0.49703	0.45662	0.46942	0.4333	0.43896	0	0.26858	0.3759
2510	5515-65	0.38755	0.32057	0.44009	0.31748	0.2881	0.30151	0.2881	0.40986	0.31119	0.53808	0.52881	0.48086	0.48901	0.45227	0.45985	0.26858	0	0.385
2511	5516-75	0.39513	0.32364	0.41939	0.32668	0.35285	0.35285	0.35285	0.40256	0.32668	0.47361	0.45877	0.44455	0.45336	0.39886	0.41226	0.3759	0.385	0
2512	5516-76	0.385	0.31119	0.41939	0.32057	0.34147	0.34147	0.33563	0.40256	0.32668	0.48597	0.47569	0.4577	0.46625	0.41346	0.41703	0.36524	0.37459	0.18861
2513	5519-64	0.48698	0.45336	0.47465	0.43784	0.44455	0.43101	0.40744	0.46837	0.42871	0.57363	0.56844	0.52412	0.53531	0.43896	0.43784	0.45445	0.40744	0.47673
2514	5519-76	0.45985	0.4333	0.45553	0.4264	0.42408	0.41939	0.37459	0.44009	0.39262	0.55436	0.54537	0.50296	0.51461	0.42756	0.4264	0.39638	0.37459	0.46625
2515	5520-74	0.43557	0.2881	0.46625	0.231	0.27404	0.26673	0.28116	0.41465	0.2394	0.53808	0.54356	0.47673	0.49703	0.43444	0.43784	0.35145	0.35564	0.37459

- Selecting the reference sample seeds: Each reference sample contains 20 seeds, which also differ from one another somewhat in genetic makeup. As a result, we needed to determine which seeds from the reference sample would serve as the reference seeds for comparison to seeds from the field samples. We calculated the Roger's distance between all seeds within each reference sample, and this produced a symmetric matrix only for reference sample seeds. The two seeds that were closest to each other in terms of genetic distance were selected to serve as the seeds to which field sample seeds would be compared. These are the two reference sample seeds (one in a row and one in a column) that are be included for comparison in Figure 3 (bright green seed IDs are reference sample seeds). All but these two seeds are dropped from the reference sample.
    - Note: As explained by the agronomist at the National Crops Resources Research Institute (NaCRRI) in Uganda, selecting the two most similar reference seeds in terms of Roger's distance is also likely to identify two seeds that are true to type for that variety. He described how these seeds that are most similar are not likely to be very dissimilar to other seeds in the reference sample. They are unlikely to be an outlier pair in terms of genetic makeup.

- Our measure of genetic authenticity is then calculated as the average Rogers Distance between the field sample seed and the two reference sample seeds of the variety that the seed is claimed to be. For example, a seed from a field sample claims to be Longe 11H. There are two Longe 11H reference seeds (Longe11H_seed1 and Longe11H_seed2). The distance between the field sample seed and Longe11H_seed1 is 0.5 and the distance between the field sample seed and Longe11H_seed2 is 0.3. The average distance would be 0.4. This is the continuous measure of genetic distance that is used as the outcome variable at the seed level. Again, lower numbers reflect a greater degree of similarity to the reference sample for the field sample's stated variety.

Appendix D: Technical Appendix on Quality Testing for Glyphosate Herbicide

Measurement of the glyphosate concentration in the field samples collected during the study was conducted using a spectrophotometric[23] method.[24] The measurement relies on the reaction of glyphosate molecules with a mixture of distilled water, ninhydrin (a chemical used to detect molecular compounds), and sodium molybdate (an inorganic sodium salt) that produces a purple color, whose intensity indicates the concentration of glyphosate in a sample. The intensity of the color is then quantified by comparing it to a known concentration (a standard) of glyphosate. This method is suitable for quality control testing (Nagaraja et al. 2011) and has previously been used for the detection of glyphosate (Bhaskara et al. 2006).

Steps used for testing:

1. A sub-sample of each bottle of the field samples of glyphosate herbicide collected was poured into 2 ml tubes for analysis.

2. A first solution of 50 ml distilled water was prepared. Distilled water is used to dilute the glyphosate herbicide, which is highly concentrated, so that testing can be performed.

3. 0.025 ml of each glyphosate analysis sample was mixed into 15 ml of distilled water (the first solution) and shaken for 30 seconds to dilute the samples. Diluted samples were needed because the type of testing used is typically conducted on very diluted samples that only have trace amounts of substances.

4. Two batches of a second solution were also prepared. For each batch, 2.5 g of sodium molybdate was weighed and dissolved in 10 ml of distilled water until it was saturated.[25] Then, 0.1 ml of this saturated solution was added to 0.5 ml of ninhydrin. This solution provides the combination of sodium molybdate and ninhydrin that reacts to produce the purple color indicating the concentration of glyphosate.

5. 0.1 ml of the diluted glyphosate solution (from step 3) was added to the saturated solution of sodium molybdate, distilled water, and ninhydrin (from step 4).

6. This mixture was then sealed in a conical tube (1.5ml) and placed in a steam water bath heater set to 100 degrees C for 2 minutes.[26]

[23] Spectrophotometry is a method to measure how much a chemical substance absorbs light by measuring the intensity of light as a beam of light passes through a sample solution.
[24] Testing was conducted by Mobile Assay (https://mobileassay.com/).
[25] A solution with solute that dissolves until it is unable to dissolve anymore, leaving the undissolved substances at the bottom.
[26] This temperature and time were determined to be optimal via testing variations for best results.

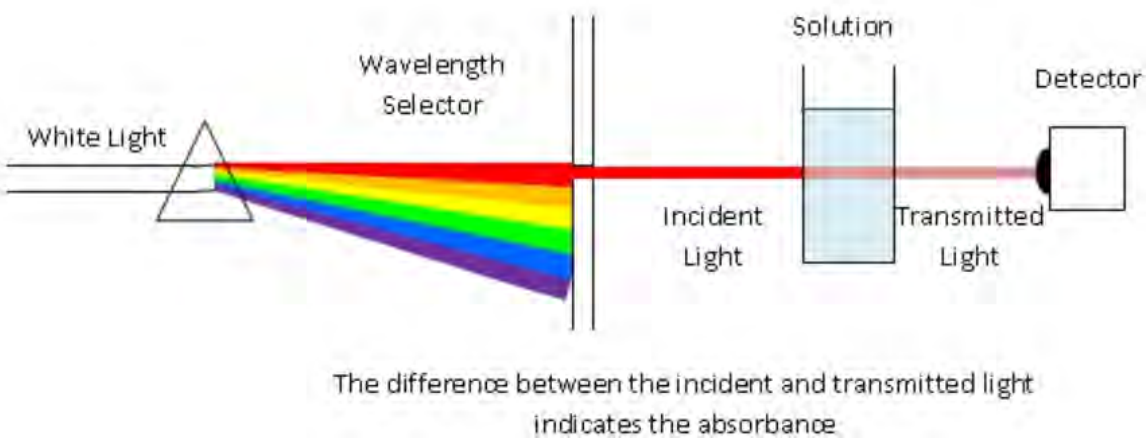
7. The heated mixture was removed, shaken, and placed in a cuvette[27] for measurement of the concentration of glyphosate present in the sample using spectroscopic absorbance measurement.

8. The mixture was then cooled to room temperature (22 degrees C, the optimal temperature for this measurement). The process of heating the solution and then cooling it to room temperature is what produces the purple color that represents the concentration of glyphosate.

9. A third solution was prepared in order to quantify the concentration of glyphosate. This third solution is the reference point (a known concentration – the standard) to which the concentration of glyphosate of the field samples is compared. 0.01 ml of 1000 ug/ml[28] glyphosate standard (known concentration of glyphosate) was added to 0.6 ml of the mixture of sodium molybdate, distilled water, and ninhydrin (the second solution). It was heated at 100 degrees C for 2 minutes, and then cooled to room temperature (22 C).

10. The concentration of glyphosate is quantified as follows, using a spectrophotometer (see Figure 1):

Figure 1.



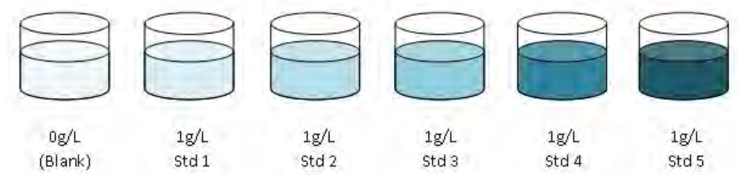
Source: https://di.uq.edu.au/community-and-alumni/sparq-ed/sparq-ed-services/spectrophotometry.

   a. The glyphosate standard, which has a known concentration of glyphosate, is diluted to different levels. The first dilution is always zero (no glyphosate), and then the amount of the standard (the concentration) in the same amount of distilled water, sodium molybdate, and ninhydrin is increased. This is done to calibrate the spectrophotometer (to provide a base to which to compare field samples). See Figure 2. The absorbance[29] of each of the dilutions is calculated using the

[27] A cuvette is a piece of laboratory equipment intended to hold samples for spectroscopic analysis.
[28] Micrograms per milliliter, where 1 microgram is equal to one millionth of a gram (10^{-6})
[29] Absorbance does not have a specific unit; it is just relative to the concentration of zero.

spectrophotometer.[30] Absorbance is the amount of light of a particular wavelength absorbed by a substance across a constant distance (light path). [31]

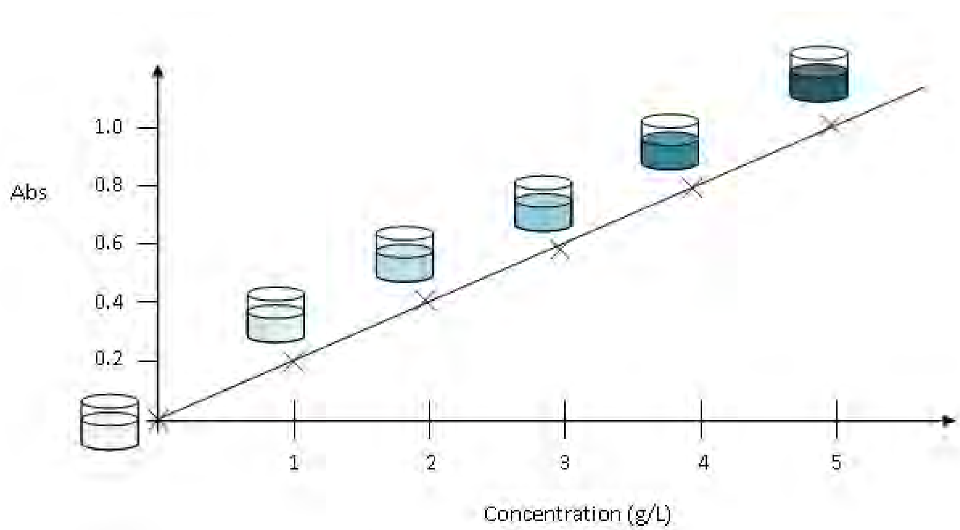
Figure 2.



Source: https://di.uq.edu.au/community-and-alumni/sparq-ed/sparq-ed-services/spectrophotometry.

b. For each level of concentration of the standard, the absorbance is plotted against the concentration. The absorbance is proportional to the concentration of that substance. A line of best fit is drawn through the points. This produces a 'standard curve'. See Figure 3.

Figure 3.



Source: https://di.uq.edu.au/community-and-alumni/sparq-ed/sparq-ed-services/spectrophotometry.

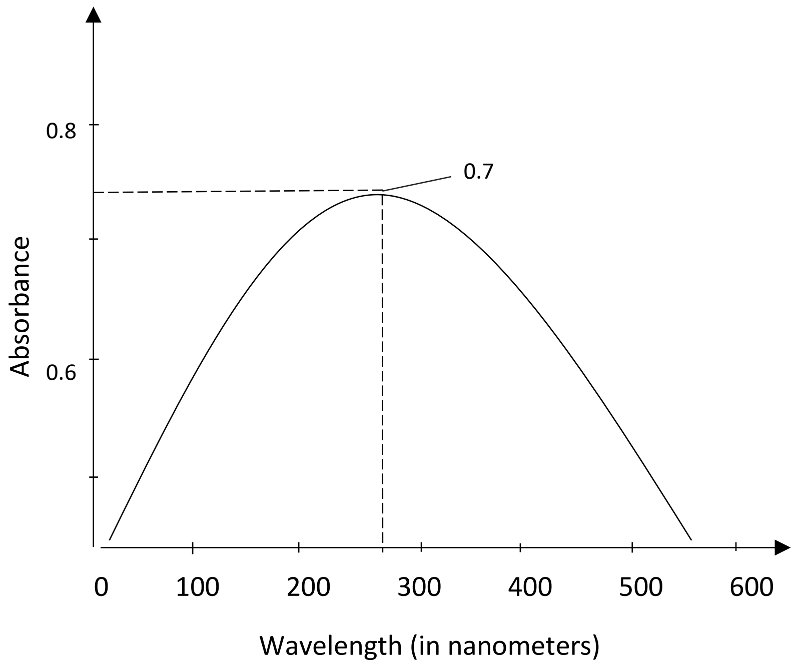
c. The diluted field sample (from step 8 – also mixed with sodium molybdate and ninhydrin) is then analyzed using the spectrophotometer. A bell-shaped curve with one peak is produced. The peak indicates the absorbance level of the field sample. See Figure 4.

---

[30] Vernier spectral analysis software and a Vernier UV-VIS Spectrophotometer were used.

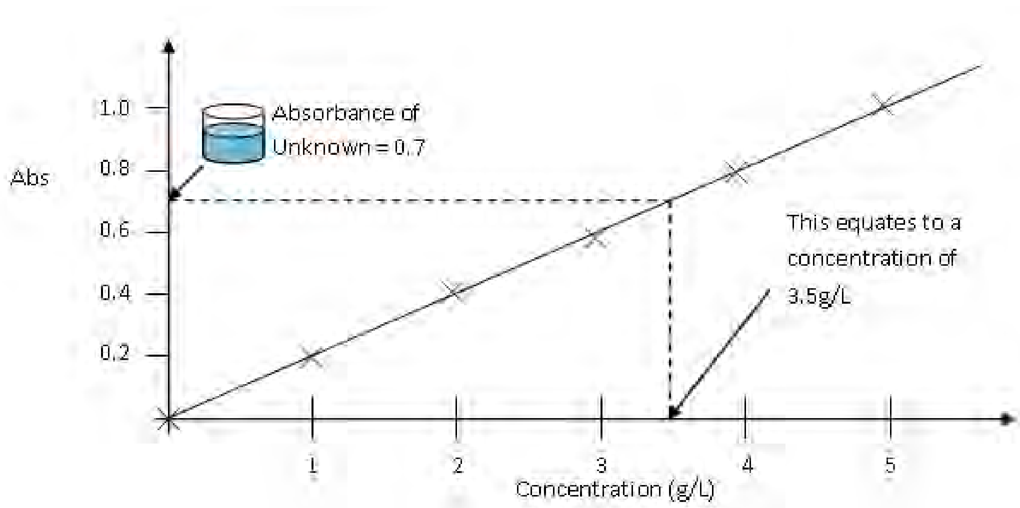
[31] 570nm is the maximum absorbance in this case.

Figure 4.



d. The absorbance level of the tested analysis sample is then plugged into the standard curve graph of the standard with known concentration to calculate the predicted concentration level of the field sample. See Figure 5. This is the outcome variable used in the analysis.

Figure 5.



Source: https://di.uq.edu.au/community-and-alumni/sparq-ed/sparq-ed-services/spectrophotometry.

11. The field samples contained other surfactants and detergents together with the glyphosate, such as polyamines and tallow amines (which are additives in glyphosate herbicide). A standard solution containing tallow amine surfactants without glyphosate was used as a negative control and it was determined that no non-specific ninhydrin reaction occurred.