Salience, Misperceptions and Trade-offs in Agricultural Technology Adoption: Experimental Evidence from Uganda

Paola Mallia*

Paris School of Economics & University of Notre Dame

November 24, 2026

Latest version: here

Abstract

Farmers often have imperfect information about the traits of the crop varieties they grow. This paper shows that this misidentification is a first-order constraint on the adoption of nutrient-enhanced and disease-resistant varieties. In a randomized controlled trial with 2,141 sweet potato farmers across 108 Ugandan villages, I compare two ways of addressing this information imperfection. The first teaches farmers to read visual cues as signals of varietal traits, flesh color for nutritional content (vitamin A) and symptoms on plants for disease. The second adds variety-specific trait information, using DNA fingerprinting to tell farmers which traits their own varieties actually carry. Learning to read the cues improves farmers' knowledge but does not change what they plant. Variety-specific trait information raises correct identification of yellow-fleshed and disease-resistant varieties and the area planted to those varieties by 25 to 59 percent. Further evidence suggests that reading the cues is not enough because farmers rarely recognize that desirable traits co-occur, and so wrongly believe these traits trade off against one another; information that reveals which varieties combine them shifts planting. These changes in adoption raise consumption of vitamin A–rich sweet potato among children under two and cut their diarrhea incidence by 52 percent, with positive spillovers to children of a second group of farmers who received only community-level DNA results. By linking objectively measured varietal adoption to child health, the results provide causal evidence that varietal misidentification limits the returns to agricultural R&D.

*E-mail: paola.mallia@psemail.eu or pmallia@nd.edu. The research was approved by the J-PAL Europe/PSE IRB and has been registered in the AEA-RCT registry (AEARCTR-0010070). I thank SPIA of CGIAR, Karen Macours, Sylvie Lambert, Kyle Emerick, Tanguy Bernard, Jack Willis, Paola Villar, Andrew Dillon, Doug Gollin, Paul Winters and participants of PSE seminars and the STEG annual conference for helpful comments and suggestions. I also thank ANR for support received through the EUR grant ANR-17-EURE-0001. This paper previously circulated under the title "Color-Blind to the Obvious: Salience, Misperception, and Technology Adoption in Uganda."

1 Introduction

Technology adoption is a central driver of agricultural productivity, yet progress in Sub-Saharan Africa has been more limited than gains seen elsewhere (Evenson & Gollin, 2003; Gollin et al., 2021; Pingali, 2012). Adoption is a complex, multidimensional decision over seeds, fertilizer, and labor (Laajaj & Macours, 2026). Farmers make these decisions under uncertainty about how inputs combine to determine returns (Conley & Udry, 2010; Foster & Rosenzweig, 1995), and may pay limited attention to signals that reveal those returns (DellaVigna, 2009; Hanna et al., 2014). These informational challenges arise not only across combinations of inputs but also within a single technology. A crop variety bundles traits such as yield, disease resistance, and vitamin content that differ in how readily farmers can observe them. Farmers may therefore hold incorrect beliefs about which variety they grow and the characteristics it possesses (Macours, 2019). Whether such misbeliefs systematically distort technology choice, leading to suboptimal adoption and lower welfare, remains an open empirical question.

This paper studies whether variety misidentification, a widespread phenomenon across crops and regions (Stevenson et al., 2023), constrains agricultural technology adoption.[1] Misperception can persist even for familiar crops. First, key traits are difficult to infer from experience: visual cues need not map directly to nutritional content, while disease resistance becomes apparent only when disease pressure is high enough to distinguish resistant from susceptible varieties, a signal that arrives erratically and can be confounded with environmental shocks (Macours, 2019). Second, informal seed systems make variety labels unreliable, so even an accurate belief about a trait can be difficult to attach to the correct local variety name (Coomes et al., 2015; Louwaars & de Boef, 2012; McGuire & Sperling, 2016). These frictions raise a central question: is teaching farmers how to recognize varietal characteristics sufficient to change what they plant, or must they also learn which specific varieties possess those characteristics?

I answer this question through a randomized controlled trial among sweet potato farmers in Uganda. Sweet potato is well suited to studying varietal information because it is widely cultivated, vegetatively propagated, and characterized by substantial varietal diversity. All varieties are grown with broadly similar practices, so switching among them requires little change in inputs or management, limiting confounds that complicate adoption decisions for input-intensive crops. At the same time, varieties

[1] I use the term to encompass two closely linked failures: farmers may not know which genetic varieties they cultivate and, critically, may not correctly identify the trait profiles of those varieties. I follow a growing literature that documents these discrepancies and refers to them as *misidentification* or *misperception* (Abay et al., 2021; Bohr et al., 2024; Mallia, 2022; Wossen et al., 2022).

differ markedly in yield, taste, storability, vitamin A content, and resistance to Sweet Potato Virus Disease (SPVD) and Alternaria Blight (AB), two major sources of production risk in Uganda. This setting allows me to study how information about varietal characteristics affects planting decisions.

The experimental design distinguishes between two information frictions that are often conflated. The first concerns *what* to look for: farmers may not notice visual cues that reveal a trait or may not know what those cues signal. In short, they may not know how to *read the cues*, a failure of noticing in the sense of Hanna et al. (2014). The second concerns *which* varieties carry a trait: even an attentive farmer may have incorrect beliefs about which of her own varieties possess it. I refer to this second friction as *misidentification*. Teaching farmers to read the cues addresses the first friction, but addresses the second only if farmers can subsequently interpret the cues correctly and the cues leave little room for error. The same flesh color can look different to different farmers, while disease symptoms can be confused with other stresses. If cue-reading is the only binding friction, teaching farmers to recognize cues should improve both knowledge and planting decisions. If misidentification also binds, such teaching may improve knowledge without changing planting; variety-specific information about which varieties actually carry the trait may also be needed to change adoption.

I study these mechanisms in a sample of 2,141 farmers across 108 villages in Uganda's four major sweet potato-producing regions. In each village, I collected leaf samples from approximately half of the sampled farmers and conducted DNA fingerprinting. Laboratory genotyping matched samples to a reference library of sweet potato varieties with documented characteristics from breeding trials. These data serve two purposes: they provide the content for cards on varietal identity and traits, and they allow me to measure adoption outcomes objectively at endline, avoiding reliance on farmer recall. Villages were randomly assigned to three arms. The control group (36 villages) received no intervention. Treatment 1 (T1) trained farmers to read visual cues as signals of varietal characteristics, using demonstrations and agronomic explanations. Facilitators showed roots of different colors and explained their nutritional implications, and used photographs to teach farmers to recognize SPVD and AB symptoms and understand disease transmission. Treatment 2 (T2) added variety-specific trait information for farmers whose crops were tested. These farmers received individual cards linking local names reported at baseline to official variety names and reporting laboratory-verified trait information for each variety they cultivated, including flesh color category and resistance status for SPVD and AB. Whereas T1 gives farmers a rule for inferring traits from what they can see, T2 provides information about the actual characteristics of their varieties. Because both treatment

arms receive the same session on visual cues, comparing T2 with T1 measures the additional value of variety-specific identification. I also randomized which household member received the information, allowing me to test whether responses depend on how much the recipient values the targeted traits.

I find that learning to read visual cues improves farmers' knowledge but does not change planting behavior, while variety-specific trait information generates large adoption shifts. Both treatments improve knowledge: farmers become 27–28% more likely to correctly report that flesh color signals nutritional value, and improve disease-symptom recognition by 0.5–1.1 standard deviations. However, only T2 improves farmers' ability to correctly identify the traits of their own varieties, raising correct identification of yellow-fleshed varieties by about 68% and of disease-resistant varieties by about 20%. These identification gains translate into greater adoption: relative to control means, T2 farmers increase the area planted with yellow-fleshed varieties by 36%, with disease-resistant varieties by 25–45%, and with varieties combining both traits by 59%. The contrast between T1 and T2 is consistent with misidentification being a binding constraint: learning to recognize cues is not sufficient to change planting when farmers do not know which varieties on their farms possess desirable characteristics.

A conceptual framework helps interpret these findings. In the model, farmers allocate land across varieties based on perceived trait values. The two information frictions described above interact with a third: beliefs that desirable traits trade off against one another. Because farmers cannot correctly identify which varieties carry which traits, they may rarely recognize that desirable attributes co-occur and may believe that nutritional value, yield or taste come at the expense of resistance to diseases, pests or drought. Baseline data are consistent with this possibility: farmers who cite a variety's nutritional value or yield often list disease, pests or drought susceptibility as its weaknesses. Under such beliefs, the framework predicts that teaching farmers to read a cue for a single trait need not increase adoption, since a farmer who comes to regard a variety as nutritious also expects it to be weaker on other dimensions. This is consistent with the absence of planting responses under T1. Variety-specific information can instead reveal which cultivars combine several advantages. Consistent with this prediction, T2 effects on area are concentrated among households that already had a DNA-verified yellow-fleshed, virus-resistant variety available on their farm at baseline, and, within this group, among those who did not know they had one: the farmers for whom variety-specific information revealed something new.

These planting shifts are accompanied by higher production and improved nutrition. T2 farmers report 87% higher harvested quantities of yellow-fleshed varieties and 50%

higher quantities of virus-resistant varieties than the control group. In households with children under two, T2 raises the probability that a child consumed vitamin A-rich sweet potato in the past seven days by 32%. T2 also reduces diarrhea incidence by 52% and the duration of the most recent episode by 1.3 days, relative to a control mean of 3 days. These health effects are consistent with a pathway in which greater production and consumption of vitamin A-rich sweet potato contributes to child health. The magnitudes are consistent with prior evidence from interventions that directly provided new orange sweet potato vines, which raised children's vitamin A intake (Hotz, Loechl, de Brauw, et al., 2012; Hotz, Loechl, Lubowa, et al., 2012) and reduced diarrhea among young children (Jones & De Brauw, 2015). I find no effects on outcomes unrelated to this proposed pathway, such as respiratory symptoms and malaria.

Within T2 villages, I also study whether information benefits households whose own crops were not tested. A second group of households received community-level summary cards describing which varieties were present in the village and their characteristics, allowing me to measure information spillovers and assess a potentially lower-cost channel. Farmers who did not receive individual DNA results but were exposed to community-level DNA results increased consumption of orange or yellow sweet potato among children under two by 37% and experienced a 61% reduction in diarrhea incidence. These findings suggest that variety-specific trait information can relax important information constraints and that its benefits can extend beyond households whose crops are tested.

This study makes three contributions to the literature on agricultural technology adoption. First, it provides what is, to my knowledge, the first experimental evidence that variety misidentification causally constrains adoption. A large literature studies information frictions about returns and required know-how (Bandiera & Rasul, 2006; Conley & Udry, 2010; Foster & Rosenzweig, 1995) and the consequences of uncertain input quality (Bold et al., 2017), while a measurement literature documents that farmers misreport the varieties they grow (Stevenson et al., 2023). I move from documenting misidentification to demonstrating its behavioral consequence: correcting farmers' beliefs about their own varieties' traits changes planting choices. These beliefs are systematic rather than merely noisy. Farmers misperceive not only whether a variety carries a trait but also how traits covary, believing that advantageous attributes trade off even though varieties combining them exist. Because farmers cannot identify which of their varieties carry which traits, they rarely observe these combinations, which can sustain mistaken beliefs about trade-offs. This helps reconcile mixed findings in the information-provision literature (Jack, 2011; Magruder, 2018): drawing attention to a trait can lower its perceived utility when farmers believe it implies a sacrifice

elsewhere, so knowledge may rise without adoption. Relative to work on learning and complementarities across inputs (Emerick et al., 2016; Laajaj & Macours, 2026), I show that informational frictions can distort choices even for a single, low-input crop.

Second, the dual use of DNA fingerprinting as both treatment and measurement is a methodological contribution. DNA fingerprinting has been used to measure adoption (Floro et al., 2018; Jaleta et al., 2020; Kosmowski et al., 2020; Maredia et al., 2016; Yirga et al., 2016); this paper uses its results to provide information to farmers as the intervention itself. DNA fingerprinting at endline also addresses a reporting bias that is especially severe in this setting: farmers misidentify the varieties they grow, and information treatments may change what they report rather than what they plant. A farmer who learns "yellow is good" may over-report yellow without replanting. My data show substantial discrepancies between reported and DNA-verified traits. By testing variety identity and trait profiles objectively and measuring cultivated area by GPS, I estimate treatment effects on actual planting. These tools are likely to become more valuable as breeding programs release varieties with increasingly complex trait bundles, for which self-reports become progressively less reliable (Wossen, Abdoulaye, et al., 2019; Wossen, Nguimkeu, et al., 2019).

Third, I provide evidence linking agricultural information, adoption, nutrition, and child health, and show that an agricultural information intervention can produce measurable health gains within a single growing cycle. Most adoption studies stop at intermediate outcomes (Foster & Rosenzweig, 2010; Magruder, 2018), while nutrition-agriculture research often lacks rigorous identification of the farm-to-health pathway (Ruel & Alderman, 2013). Using objectively measured variety choice, I find that T2 changes planting, increases reported production of targeted varieties, raises children's consumption of vitamin A-rich sweet potato, and reduces diarrhea among children under two. This pattern is consistent with the established protective effect of vitamin A against child morbidity (Imdad et al., 2022; Mayo-Wilson et al., 2011), and the absence of effects on unrelated symptoms is consistent with this interpretation. The findings add causal evidence on the health impacts of agricultural interventions (Bouis & Saltzman, 2017; de Brauw et al., 2018; Hotz, Loechl, de Brauw, et al., 2012; Hotz, Loechl, Lubowa, et al., 2012; Jones & De Brauw, 2015).

The findings have concise policy implications. Misidentification can erode returns to investments in breeding, extension, and seed systems, while correcting it can unlock latent demand for existing high-performing varieties. Agricultural extension can complement training on how to recognize trait cues with information that identifies which traits farmers' own varieties actually possess, especially where farmers perceive trade-offs among traits. DNA fingerprinting of subsamples or observable agronomic proxies may provide practical ways to supply this information. Because

community-level information also benefits households whose crops were not tested, diagnostic information may not need to be collected for every household. The observed nutrition and health outcomes also point to potential gains from coordination between agricultural and public-health programs.

The remainder of the paper is organized as follows. Section 2 provides background on sweet potato production in Uganda and documents the extent of varietal misidentification. Section 3 develops the conceptual framework and predictions, with the full model and its derivations in Appendix A. Section 4 describes the data and experimental design. Section 5 presents the main results on cue reading, varietal identification, adoption, production, consumption, health, and spillovers. Section 6 analyzes perceived trade-offs in varietal traits. Section 7 examines heterogeneous treatment effects by baseline information intensity and respondent gender. Section 8 reports robustness analyses. Section 9 concludes; Appendix F examines the costs of scaling the intervention.

2 Context

2.1 Sweet Potato Production in Uganda

Sweet potato is a staple crop throughout Uganda and is crucial for household food security and rural livelihoods. In 2020, the country produced roughly 1.2 million metric tonnes from an estimated 466,000 hectares ((UBOS), 2020). Cultivation occurs across all four major regions, with most smallholder farmers relying on the crop primarily for own consumption. Varietal diversity is substantial at baseline: households grow an average of 3.31 varieties and up to 12, reflecting the fact that different varieties embody distinct agronomic and consumption-related traits. Plausibly, farmers choose varietal portfolios that balance stress resilience, food quality, and complementary uses.

Two complementary breeding objectives have shaped sweet potato R&D efforts in East Africa. The first concerns production risk. Sweet potato virus disease (SPVD) and Alternaria Blight (AB) substantially reduce yields, often by 50%[2]. Breeding programs have responded by developing resistant varieties, which are disseminated through both formal multiplication channels and informal vine exchange.

The second objective concerns nutrition. In 2011, about one in three Ugandan children aged 6–59 months were vitamin A deficient (World Health Organization, 2011). Unlike caloric malnutrition, micronutrient deficiencies often manifest without visible symptoms, making them difficult to detect and address through dietary changes alone

[2]In my sample, farmers report roughly 50% harvest losses when they observe symptoms of SPVD or AB on their plots at baseline.

(Bouis & Saltzman, 2017). Vitamin A is critical for maintaining mucosal integrity and supporting immune function (Stephensen, 2001). Meta-analyses show that vitamin A supplementation reduces diarrhea-related morbidity and all-cause mortality among young children (Mayo-Wilson et al., 2011), although short-run effects on malaria, respiratory infections, or measles are less consistent (Imdad et al., 2022).

Together, these objectives have led breeding programs to target varieties that combine multiple traits. Biofortification initiatives have focused on orange-fleshed sweet potato (OFSP), which contains high beta-carotene levels (Low et al., 2007). HarvestPlus has led large-scale dissemination campaigns since the early 2000s (HarvestPlus, 2012). Yellow-fleshed varieties (YFSP), although never formally promoted, also contain meaningful beta-carotene. A representative yellow variety provides approximately 110 μg retinol equivalents per 100g after bioconversion (Rodriguez-Amaya & Kimura, 2004); thus, a 6–12-month-old child consuming 200–230g per day would receive nearly half of the recommended vitamin A intake from a single source.

Despite these breeding efforts, the varietal landscape exhibits wide heterogeneity in adoption. Table 1 summarizes the distribution of nutritional and resistance traits based on DNA-verified leaf samples from 1,000 households at baseline. Adoption of OFSP remains low: only 6.2% of farmers cultivate at least one orange-fleshed variety, despite more than a decade of promotion.[3] In contrast, 30.1% of farmers grow yellow-fleshed varieties and 35% grow white-fleshed varieties. Although white-fleshed varieties contain no β-carotene (Tumwegamire et al., 2011), they are long-standing traditional types that remain widespread in the landscape and are valued for their taste (Mwanga et al., 2017, 2021).

Resistance traits are far more prevalent than nutritional traits. At baseline, 61.7% of households cultivate at least one SPVD-resistant variety, and 54.3% grow an AB-resistant variety. Importantly, this does not imply a scarcity of *"win-win"* varieties: varieties that combine higher nutritional content with virus resistance are already accessible to many households. Panel A of Table 1 shows that 22.2% of farmers grow at least one variety that is both yellow-fleshed and virus-resistant. The regional breakdown (Panel C) indicates that these varieties are present in all four regions, ranging from 6.1% of households in the Eastern region to 47.8% in the Central region. Nevertheless, spatial heterogeneity remains substantial, as shown by Figures 1–2.

[3]This aligns with findings from the Uganda Harmonized Integrated Survey 2021/22, which reports even lower rates (3%) of objectively measured adoption.

2.2 Do Farmers Know What They Grow?

The presence of multiple varieties does not necessarily mean that farmers correctly identify their traits. Whether they can is central to understanding observed adoption patterns, and their ability to learn differs sharply across trait types. Flesh color is directly observable at harvest and can, in principle, support learning about nutritional content. Even so, learning may fail if farmers do not understand that flesh color signals nutrition rather than other attributes. Disease resistance is much harder to infer from experience. It becomes visible only under sufficiently high disease pressure and often resembles other agronomic stresses such as drought or soil nutrient deficiency. Symptoms of SPVD in particular are frequently mistaken for nitrogen deficiency or moisture stress; only 17% of farmers in my sample correctly attribute these symptoms to disease. Disease pressure also varies across seasons and locations due to changes in vector prevalence, weather, and infection dynamics. As a result, farmers observe noisy and sometimes misleading performance signals: a resistant variety may appear susceptible in a low-pressure year, while a susceptible variety may perform well by chance. Even farmers growing multiple varieties side by side may fail to learn true resistance status when disease pressure is weak or confounded by other stresses. These conditions make systematic misidentification likely.

To measure the extent of misidentification, I compare farmers' self-reported varietal identification with DNA-fingerprinting results. Table 2 presents this comparison using endline data from control-group households. The results show large gaps between farmers' beliefs and the objective characteristics of the vines on their fields. For flesh color, misclassification is limited for orange-fleshed varieties, mainly because adoption is rare. Orange flesh is correctly identified in only 2.7% of farmer–variety observations, while 4.5% are non-orange varieties classified as orange. False negatives (0.9%) are uncommon, and true negatives are high (91.8%). In contrast, misclassification is much more common for yellow-fleshed varieties, despite flesh color being directly observable. Farmers correctly identify a yellow-fleshed variety in 14.2% of observations, attribute yellow flesh to a non-yellow variety in 28.4%, and fail to recognize a yellow-fleshed variety in 26.1%. Only 31.1% of observations are non-yellow varieties correctly reported as such. One reason is that flesh color is less distinct than it appears. Yellow-fleshed varieties are typically pale yellow, close to cream or white (Figure A1), and the intensity of color in a given root can vary with growing conditions and root maturity, so the same variety may look white in one harvest and yellow in another. Together, these patterns suggest substantial difficulty mapping an observable trait onto varietal categories.

Misclassification is similarly severe for resistance traits. For SPVD resistance, 62.2% of observations are resistant varieties correctly identified as such, but in 24.1%

farmers fail to recognize resistance that is present. False positives are also common (10.8%), and true negatives are rare (2.8%). AB resistance shows a comparable pattern: farmers correctly identify resistant varieties in 44.4% of cases, while both false positives (27.7%) and false negatives (22.3%) are substantial.

These errors are reinforced by how farmers interpret performance signals. Farmers who believe a variety is unlikely to be disease-resistant may not attend to evidence that contradicts this belief. When a variety performs well under disease pressure, farmers may attribute this to favorable plot conditions rather than genetic resistance. Incorrect priors, often shaped by older varietal generations with stronger trade-offs, can therefore persist.

These belief patterns help explain why farmers may not shift resources toward varieties that combine multiple desirable traits, even when such varieties are locally available at zero cost. In particular, they imply a perceived trade-off between traits. To assess this, I asked control-group farmers to list up to three "good" traits and three "bad" traits for each cultivated variety. Figure 3 shows the conditional probability of reporting a given bad trait given a reported good trait.[4] Farmers report strong trade-offs. Those who cite high nutrition as a positive trait often list disease susceptibility (18.8%), pest susceptibility (11.5%), or drought vulnerability (18.8%) as weaknesses. Similarly, farmers who emphasize pest, drought or disease resistance often report bad cooking properties or bad taste. Together, these responses point to a mental model in which gains along one dimension are believed to require losses along others.

However, this view contrasts with objective trait bundles. As shown in Table 1, 22.2% of farmers cultivate varieties that combine high nutritional content and virus resistance based on DNA fingerprinting, despite the fact that farmers believe such combinations are rare. This gap likely reflects both the difficulty of learning about resistance from field experience and the persistence of beliefs formed under earlier breeding regimes or traditional varieties with stronger trade-offs.

Finally, farmers' self-reported confidence suggests that these patterns do not mainly reflect uncertainty. Figure 4 shows the distribution of confidence levels across variety traits. For flesh color, farmers express overwhelming certainty: 98% report high confidence (levels 4–5) for orange varieties and 94% for yellow varieties, with approximately three-quarters reporting absolute certainty (level 5). Even for disease-resistance traits, which are not directly observable, roughly two-thirds of farmers report high confidence. This combination of high confidence and low accuracy points to systematic misbelief rather than acknowledged uncertainty, suggesting farmers are making land allocation decisions based on incorrect but strongly held beliefs about variety characteristics.

[4]The figure uses endline data from the control group.

3 Conceptual Framework

The previous section showed that many farmers misidentify the traits of the varieties they grow and perceive strong trade-offs between nutrition, resistance, and other attributes, even though, according to the genetic information, varieties combining nutrition and resistance are common. This section develops a simple framework that formalizes these patterns and links them to farmers' land allocation decisions under the experimental treatments. Farmers allocate land across varieties with multiple agronomic and nutritional traits, and form beliefs about those traits from the visual cues their fields provide. I focus on three informational frictions that can distort these choices: (i) not knowing how to read the cues that signal the targeted traits, (ii) misidentification of which varieties carry those traits, and (iii) mistaken beliefs about whether desirable traits can co-occur. These frictions map into observable outcomes in the experiment (recall of the cue mapping, the accuracy of the traits farmers report for their own varieties, and the land they allocate to each variety) and generate predictions that guide my empirical analysis. Appendix A sets out the full model and derives each prediction.

3.1 Setup

Each season, farmer i decides how much land $a_i(v)$ to give each variety v she holds. Farmers do not value traits for their own sake. They care about what a variety delivers: how well it nourishes the household, how much it yields and how reliably, how it tastes and stores. These preferences over outcomes are fixed throughout. The intervention targets two traits, indexed by k: the vitamin A content of the root, which flesh color signals, and resistance to disease. I write $x_k(v)$ for the level of trait k in variety v.

Farmers do not observe $x_k(v)$ directly. For each targeted trait they observe a visual cue in their own field, $s_{ik}(v)$, which equals the trait plus noise: flesh color at harvest for nutritional content, and the appearance of the plant under disease pressure for resistance. Both cues are informative but noisy: yellow flesh does signal beta-carotene but can be pale and read as white, and resistance reveals itself only when disease pressure is high and even then can be easily confounded with drought or nutrient stress. Farmer i's belief about trait k in variety v combines a prior with the cue:

$$\tilde{x}_{ik}(v) = (1 - \rho_{ik})\, \tilde{x}^0_{ik}(v) + \rho_{ik}\, s_{ik}(v), \qquad \rho_{ik} \in [0, 1]. \tag{1}$$

The prior $\tilde{x}^0_{ik}(v)$ collects what the farmer believes from local names, past experience, and hearsay. The weight ρ_{ik} captures how much the farmer uses the cue: whether she notices flesh color and treats it as a signal of nutritional content, or plant symptoms

as a signal of disease.[5]

The value farmer i perceives in a unit of land planted with variety v is

$$\tilde{\pi}_i(v) = \bar{\pi} + \sum_k \big(\omega_{ik} - \tau_i\,\kappa_{ik}\big)\,\tilde{x}_{ik}(v) + \eta_i(v), \tag{2}$$

where $\bar{\pi}$ is the value of a reference variety, $\omega_{ik} \geq 0$ is the value to her of the outcomes that trait k delivers (nutrition for vitamin A content, the harvest for disease resistance), $\tau_i\,\kappa_{ik} \geq 0$ is the cost she believes the trait carries for other outcomes, such as taste or yield, and $\eta_i(v)$ collects the value of everything else about the variety. The return to a variety falls as it is given more land: for example, the first plots given to a variety are those where it does best, planting material and labor at planting are limited, and a household that eats what it grows has a taste for variety in its diet and reason to spread risk. The farmer therefore gives each variety land in proportion to the value she perceives in it,

$$a_i(v) = \frac{\tilde{\pi}_i(v)}{\phi_i}, \tag{3}$$

where $\phi_i > 0$ governs how fast the return falls. She holds several varieties, gives more land to those she values more, and changes the land on a variety in proportion to the change in its perceived value. With correct beliefs and no perceived trade-offs, this is the allocation she would choose with full information; the frictions below drive a wedge between the two.

Adjustment within a season. Three features of the setting are likely to shape how these responses show up in a single season. First, I treat the problem as separable across varieties, so that the land for a variety that gains in perceived value can come from other crops or fallow rather than from the farmer's other sweet potato varieties. Second, (3) describes the allocation the farmer moves toward, and within a single season the adjustment may show mainly as expansions: sweet potato vines cannot be stored and a variety is conserved only by planting it, so a farmer who has just revised her view of her varieties may prefer to keep all of them in the ground until she has seen how they perform. Third, a variety she does not hold generally has to be sourced from other farmers through informal exchange, which is costly. With information arriving shortly before planting, responses are therefore more likely to appear first in the land given to varieties farmers already hold than in the set of varieties they hold.

[5] I do not impose Bayesian updating. A Bayesian farmer would weight the cue according to its precision, and a session that leaves the cue unchanged could not move that weight. Here the session can, and the weight may end up above what the precision of the cue warrants.

3.2 Informational Frictions

Three frictions can keep a farmer from the allocation she would choose with full information. The first two concern her beliefs about which varieties carry a trait; the third concerns her belief that a desirable trait comes at the expense of something else.

Friction 1: Not Knowing How to Read the Cues

A farmer may not notice the cue in her field, or may see it without knowing what it signals, and so not use it to form beliefs: $\rho_{ik} \approx 0$ in (1). As in Hanna et al. (2014), the farmer observes the color of her roots but does not treat it as information about nutrition, so her belief rests on the prior alone.

Friction 2: Misidentification of Which Varieties Have Which Traits

A farmer who reads the cue may still hold incorrect beliefs about which of her varieties carry it, $\tilde{x}_{ik}(v) \neq x_k(v)$. Equation (1) makes the two sources explicit. The prior may be wrong, which matters when ρ_{ik} is low. And the cue may be noisy: even a cue she relies on fully ($\rho_{ik} = 1$) can leave the belief far from the truth, and may move it in the wrong direction. A variety is undervalued if it possesses desirable traits the farmer does not recognize, and overvalued if desirable traits are incorrectly attributed to it.[6] Empirically, I measure misidentification by comparing farmers' reports $\tilde{x}_{ik}(v)$ to DNA-based measures $x_k(v)$, which is the basis for Prediction 2 and the true-positive outcomes in Section 5.2.[7]

Friction 3: Perceived Trait Trade-offs

Farmers may believe that a variety that is good on one dimension must be weak on another: that a nutritious variety is more fragile, or that a resistant one tastes worse. Section 2 documents such beliefs (Figure 3). The parameter $\tau_i \in [0, 1]$ in (2) measures how strongly farmer i holds this view; with $\tau_i = 0$ she perceives no trade-offs. It enters her choices in two ways.

First, it lowers the value she perceives in a trait. She associates the trait with costs for outcomes about which the experiment provides no information, such as taste,

[6] The notation treats variety identity as given to focus on trait-level beliefs, which are the decision-relevant information. In practice, farmers may not know official variety names, but this matters only insofar as it prevents them from learning traits. The DNA intervention solves both problems simultaneously by linking local variety names to documented trait profiles.

[7] For flesh color, the report is the color the farmer gives for the variety. For a farmer who uses the cue, reporting that a variety is yellow-fleshed and believing that it carries vitamin A are the same statement, so I treat the reported color as her belief about the trait.

cooking quality, or yield. This is the term $\tau_i \kappa_{ik}$ in (2): the value she perceives in trait k is $\omega_{ik} - \tau_i \kappa_{ik}$, not ω_{ik}.

Second, it can shape what she believes about a trait she has not been told about. A farmer who thinks that nutrition and resistance trade off may expect a variety she regards as rich in vitamin A to be less resistant, and the reverse. For simplicity I write this inference as running from flesh color to resistance; the same logic applies in the other direction and leads to the same predictions.

I use the same τ_i in both because these beliefs reflect one view: evidence against one trade-off is a reason to doubt the others. As a result, when a farmer starts to see a variety as rich in vitamin A, its value to her does not necessarily rise: she gains nutrition but, in her mind, gives up some resistance, taste, or yield.

Beliefs versus preferences. Preferences over outcomes are fixed, and the channels through which the treatments act are beliefs: about which varieties carry a trait ($\tilde{x}_{ik}$) and about what comes with it (τ_i). The session common to both arms also explained why vitamin A and disease resistance matter. If this led farmers to place more value on the targeted traits, the change is common to both arms and does not affect the comparison between them. Because I randomize the gender of the respondent who receives the information (Section 4.5), I can test whether the response to the same information depends on how much the recipient values these outcomes (Section 7.2).

3.3 Treatments and Predictions

I now map the treatments into the framework and state the predictions. Appendix A derives each of them.

Treatment 1: visual cues. T1 teaches farmers to read flesh color as a signal of nutritional content and plant symptoms as a signal of disease, and explains why each trait matters. In terms of the framework, it raises the weight ρ_{ik} that farmers place on the cues. It does not change the cues themselves, it provides no variety-specific information, and nothing in it contradicts the farmer's view of trade-offs, so τ_i is unchanged. Beliefs therefore move toward the cue as the farmer judges it. They become more accurate only if that cue is closer to the truth than the prior it replaces.

Treatment 2: visual cues plus variety-specific trait information. T2 delivers the same session as T1 and adds DNA-based identification of each sampled farmer's varieties and their traits. Because the session is identical across arms, the shift in the weight on the cues is the same, an assumption of the design that I test empirically

in Section 5.1. The individual card states the traits of each DNA-matched variety of a sampled farmer, so that for these varieties $\tilde{x}_{ik}(v) = x_k(v)$, whatever the weight on the cue or its noise. The card replaces both the farmer's reading of the cues and any inference she draws from one trait about the other. This is the formal difference between the two treatments. T1 changes the weight a farmer places on a noisy signal; T2 replaces the signal.

The cards can also weaken the farmer's view of trade-offs. A card showing a variety that combines vitamin A content and disease resistance contradicts that view directly, and a farmer who sees one trade-off fail has reason to doubt the others, so τ_i falls for a farmer shown such a variety. I expect the revision to be larger when the variety is on her individual card, since it is a plant she grows, than when it appears only on the community card, which lists the varieties of the sampled farmers in her village. Community-information households receive only the community card. They learn which varieties in their village carry the targeted traits, which they can apply to their own fields to the extent that they recognize those varieties among the ones they grow, and they see the same evidence against trade-offs.

Comparing T2 with T1 therefore measures the additional effect of variety-specific trait information on top of the session, holding the weight on the cues fixed.

Prediction 1 (Knowledge)**.** *The session increases knowledge of how to read the visual cues and map them to traits. Both arms receive the same session, so this holds for T1 and T2 alike (Section 5.1).*

Prediction 2 (Trait identification)**.** *T2 makes farmers' beliefs about which of their varieties carry the targeted traits more accurate, as measured against the DNA results. T1 does so only if the cues, as farmers read them, are closer to the truth than the beliefs they held before (Section 5.2).*

Prediction 3 (Reading the cues and land)**.** *Under T1, farmers reallocate land according to the traits they believe their varieties have, which need not be the traits the DNA results show. The response is small, or even negative, when a farmer believes that a targeted trait comes at the expense of resistance, taste, or yield. Within the season this shows as no expansion toward the varieties she believes carry the targeted traits (Sections 5.3 and 6).*

Prediction 4 (Reallocation toward varieties with targeted traits)**.** *Under T2, varieties whose targeted traits the farmer had not recognized gain land within the season. Varieties she had overvalued are expected to lose land only in later seasons, once she has seen how the varieties on her card perform. The response is concentrated among farmers who hold a variety that combines the two targeted traits, especially those who*

did not know it: the card corrects both beliefs at once and can weaken the view that desirable traits trade off (Sections 5.3 and 6.2).

Prediction 5 (Total sweet potato area)**.** *Because expansions are likely to come before reductions, T2 is expected to raise total sweet potato area within the season, as well as the area under the varieties with targeted traits (Section 5.3).*

4 Data and Experimental Design

This section describes the data and experimental design used to evaluate how information provision affects farmers' knowledge, perceptions, and adoption of sweet potato varieties with nutritional and agronomic traits. The study combines a randomized information intervention with DNA fingerprinting analysis, which serves dual purposes: delivering variety-specific trait information to farmers as part of the treatment, and providing objective measurement of varietal adoption independent of self-reports.

The section proceeds as follows. Section 4.1 describes the construction of the DNA reference library used to identify sweet potato varieties and trait-level characteristics. Section 4.2 details the sampling strategy. Section 4.3 provides an overview of the baseline and endline surveys, including details on timeline of activities. Section 4.4 details the experimental design. Section 4.5 outlines the measurement innovations employed in this study. Section 4.6 examines attrition, compliance, and baseline balance. Section 4.7 sets out the estimating equations.

4.1 Reference Library and DNA-Fingerprinting Analysis

A key empirical challenge in studying agricultural technology adoption is the pervasive misidentification of crop varieties and associated traits by farmers (Stevenson et al., 2023), which makes self-reported adoption measures unreliable. To overcome this challenge, this study employs DNA fingerprinting as a core component of the empirical strategy. DNA fingerprinting refers to the crop varietal identification of plant samples to establish their precise genetic identity. A reference library, a curated database of genetic profiles for all released and commonly grown varieties, is used to match field samples to known varieties. In practice, the reference library functions as a benchmark against which unknown samples collected from farmers' fields can be compared, enabling objective identification of the varieties cultivated and the traits they carry. This approach allows both the delivery of trait-specific information and the measurement of actual, rather than perceived, adoption behaviors.

Field Collection Protocol. During data collection, each farmer accompanies the enumerator to their sweet potato plots. On the plot, the farmer identifies each variety they are growing, and the enumerator collects a leaf sample from each variety. Samples are stored in sterile, labeled containers and sent to the laboratory for genotyping. The resulting genetic profiles are then matched against the reference library to determine the official variety name and associated traits.

Construction of the Reference Library. The DNA reference library was constructed using genotypic data from 133 sweet potato varieties compiled from international agricultural research centers, national breeding programs, and local germplasm repositories (same as Uganda National Panel Survey). The sample spans both improved and local landrace varieties, and reflects the genetic diversity present in the study regions. All varieties were genotyped using standardized protocols to ensure cross-source comparability and high matching precision[8].

Varietal Composition. Among the 133 varieties in the reference library, 86 have documented data on flesh color, a key phenotypic trait for the study's biofortification focus. To address missing values, I apply a cluster-based imputation method,[9] enabling classification of all 133 varieties by the presence or absence of orange flesh, and 102 varieties by the presence or absence of yellow flesh. These binary classifications facilitate linking varietal identity to potential vitamin A content, even in the absence of complete phenotypic profiles. Virus resistance traits, specifically resistance to Sweet Potato Virus Disease (SPVD) and Alternaria Blight (AB), the two most common sweet potato diseases in Uganda, were recorded for 90 varieties and imputed for 109, ensuring broad trait coverage across the reference set.

Sample Matching Results. A total of 3,316 leaf samples were collected from farmer plots during the baseline (March 2022) and processed for genotyping. Of these, 1,641 samples (49 percent) were successfully matched to a variety in the reference library. The remaining 51 percent of unmatched samples likely reflect local landraces or uncharacterized cultivars absent from formal breeding programs (the majority of which are white or purple fleshed). All matched samples carry trait information on orange flesh color; 1,210 also include yellow-flesh classification, and 1,264 matched

[8]See Ilukor et al. (2025); Poets et al. (2020) for more details on the DNA-fingerprinting methodology.

[9]Varieties with missing trait data were assigned traits based on cluster membership, where clusters were formed from genetically nearly identical varieties. This approach assumes that flesh color and virus resistance, highly heritable qualitative traits primarily under genetic control, are shared among genetically identical varieties.

samples have documented virus resistance attributes.

Role of DNA in the Experimental Design. The DNA data serve both an intervention and an evaluation function. At baseline, the matched varietal identities were used to generate individual and community cards for farmers, including verified variety names and trait attributes. This diagnostic component provided farmers with information they would not have been able to observe or infer on their own. At endline, new leaf samples were collected and processed using the same protocol, allowing me to track changes in the varietal composition of plots before and after treatment. This enables precise measurement of trait-specific adoption (e.g., biofortification or disease resistance) at the farmer- and plot-level, across treatment arms.

Validation and Measurement Value. To further demonstrate the value of the DNA-based measurement strategy, I compare trait information derived from DNA with farmers' own reports of varietal characteristics. Figure 5 displays the relationship between objectively measured sweet potato flesh color (x-axis) and farmers' self-reported color at baseline (represented by color-coded segments within each bar). The x-axis categorizes sweet potato varieties based on their true color as determined by a reference library (e.g., White, Yellow or Orange). For each objective color group, the corresponding bar is segmented by the share of varieties that farmers classified into different self-reported color categories: SR White (white), SR Yellow (yellow), SR Orange (orange), and SR Purple (purple). The y-axis shows the percentage of varieties falling into each self-reported category, summing to 100% within each objective color. The figure highlights systematic discrepancies between objective and perceived color. For instance, varieties with objectively orange flesh are often under-reported as orange and instead reported as yellow or even white. Yellow varieties also exhibit significant misreporting: many are reported as white. These discrepancies validate the central premise of the study: that farmers often do not accurately perceive varietal traits, and that objective measurement is essential for meaningful evaluation of trait adoption[10].

Limitations. While the DNA reference library provides a powerful tool for varietal identification, some limitations should be noted. First, not all varieties grown by farmers are necessarily captured in the library, particularly undocumented landraces. However, these can be excluded to be orange fleshed varieties, while it may still be the case that there is some disease-resistance or yellow-fleshed. Second, the use

[10]Due to the lack of an appropriate measure of self-reported disease resistance for each variety at baseline, I cannot perform the same comparison exercise for SPVD and AB using the self-reported and DNA-fingerprinting measures. However, only 17 percent of farmers correctly classify disease symptoms, while the rest attribute them to climate-related shocks.

of imputed trait classifications introduces some uncertainty in the phenotype-trait mapping. However, this is likely to be negligible as this would be due to mutations happening in the field. While I do not capture the full trait spectrum of all possible varieties, my results reflect the impact of the intervention on the varieties mentioned in the treatment, which constitute the majority of those disease resistant or nutritionally relevant. As such, the analysis provides a proof of concept for how access to variety-specific trait information can influence farmers' varietal decisions.

4.2 Sample

The sample comprises households from 108 sweet potato-producing villages located across 54 subcounties in Uganda's four major regions: Central, Eastern, Northern, and Western. These villages were originally surveyed in 2011 under the CGIAR's Diffusion and Impact of Improved Varieties in Africa (DIIVA) project, which provides the sampling frame for this study. The sample is representative of major sweet potato-producing areas and spans the country's main agroecological zones.

For the baseline of this study in 2022, I surveyed 2,150 households, a target of 20 per village across the 108 villages. After excluding nine that did not consent to participate, my baseline analysis sample comprises 2,141 households.

As part of both baseline and endline data collection, DNA fingerprinting analysis was conducted on sweet potato leaf samples collected from a randomly selected subsample of up to 10 households per village; in total, 1,000 households provided usable leaf samples. These data form the basis for both treatment assignment and the objective measurement of varietal adoption, and they constitute a unique panel dataset.

The remaining households in each village were surveyed without DNA fingerprinting, enabling analysis of the effects of receiving varietal information not about one's own variety, but the varieties of the other farmers in the same village. This design allows me to assess whether less costly, non-individualized information can still influence consumption and health outcomes, providing insight into the cost-effectiveness of alternative information delivery strategies (as in Corral et al. (2020)).

4.3 Baseline and Endline Data Collection

The data collection activities for this study involved two main rounds: a baseline survey conducted between February and March 2022 and an endline survey conducted between February and March 2023.

Both surveys administered a structured household questionnaire, consisting of the following core modules:

- Household Demographics: Collected detailed information on household composition and education.
- Agricultural Production: Focused on sweet potato cultivation during the *first season* (March–September) and *second season* (October–March). This module covered area planted, input use, variety names and characteristics, sources of planting material, yields, and market participation. Farmers were also asked about their exposure to sweet potato dissemination campaigns, extension services, training received, and varietal knowledge.
- Dietary Recall: Collected food consumption data for two children per household (one aged 0–5 and one aged 6–10, or up to age 15 if no younger children were present) and for the youngest adult male and youngest adult female in the household, prioritizing pregnant women when applicable.

The endline survey expanded upon the baseline by incorporating a health module that collected data on up to two children aged 0–5 years and two children aged 6–10 years per household. This module recorded respondent-reported illness episodes over the past two weeks and six months, enabling analysis of potential health impacts associated with the intervention.[11]

In both rounds, objective varietal identification was conducted using DNA fingerprinting on sweet potato plots cultivated during the second season. Following the protocol described in Section 4.1, enumerators collected leaf samples from up to five distinct varieties per plot and recorded GPS measurements of plot boundaries. This approach enables precise tracking of varietal adoption and misidentification patterns over time and serves as a benchmark for validating self-reported varietal information.

4.4 Experimental Design

The experimental design exploits the panel structure of the genetic identification data to evaluate how access to trait-specific information affects farmers' varietal knowledge and adoption. Laboratory analysis of the baseline DNA subsample described in Section 4.2 determined the genetic identity of the varieties, which in turn revealed two key varietal traits: (i) flesh color of the sweet potato root, linked to vitamin A content, and (ii) resistance to Sweet Potato Virus Disease (SPVD) and Alternaria Blight (AB).

[11] In addition, height measurements were collected for children under five for another quasi-experimental study (Macours et al. (n.d.)). However, given the short time frame between the intervention and endline data collection, approximately six months, I do not expect measurable changes in child growth outcomes over this period and therefore do not use these data in the main analysis.

Following the baseline DNA analysis, villages were stratified by region and by the proportion of sampled varieties with orange flesh and/or disease resistance. Villages were then randomly assigned to one of three study arms: two treatment groups and one control group, with 36 villages per arm.

Treatment Arm 1 (T1) involved a session teaching farmers to read visual cues as signals of two traits: (i) the nutritional value (Vitamin A content) of sweet potato varieties based on flesh color, and (ii) the identification and management of two prevalent sweet potato diseases, Sweet Potato Virus Disease (SPVD) and Alternaria Blight (AB).

Facilitators used visual aids to explain how flesh color correlates with vitamin A content, highlighting that orange-fleshed varieties contain the highest levels, yellow-fleshed varieties moderate levels, and white or purple-fleshed varieties little to none. The sessions emphasized the health importance of vitamin A, particularly for children and pregnant women, focusing on its general role in immune function and vision. However, they did not reference the specific short-term symptoms that vitamin A supplementation may alleviate, such as diarrhea.

The plant disease management component used visual materials to illustrate common symptoms of SPVD and AB (Figures A2 and A3). Facilitators explained that these diseases weaken plants and increase susceptibility to further infections. Farmers were informed that while insecticides can eliminate vectors, they cannot cure already infected plants. Timely removal of diseased plants was promoted as a mitigation strategy.

Notably, Treatment Arm 1 did not provide any information derived from the DNA fingerprinting analysis.[12]

Treatment Arm 2 (T2) delivered the same session as Treatment Arm 1 and added variety-specific trait information for each sampled farmer's varieties, obtained through DNA fingerprinting. Within Treatment Arm 2 villages, farmers received different information depending on whether they had been randomly selected for DNA sample collection at baseline:

- **Individual-information households:** The 10 farmers per village whose leaf samples had been collected received individual cards containing:
  - the farmer's name, village, and location;
  - the list of varieties cultivated on their farm using local names;
  - the official names of those varieties (where known);
  - whether each variety was resistant to SPVD or AB;

[12]See Appendix B more details on the protocols and visual aids.

– the flesh color of each variety, as an indicator of vitamin A content.

Each card included color photographs of disease symptoms and healthy vines to reinforce visual learning. An example is shown in Figure A4. In addition to their individual cards, these households received a community-level summary card listing all varieties identified across sampled farmers in their village. The provision of both individual and community-level information was designed to encourage information sharing and the informal exchange of planting material among farmers.

- **Community-information households:** The remaining 10 farmers per village who were not selected for DNA fingerprinting received only the community-level summary card. This card listed the varieties grown by the individual-information households (identified by farmer name) along with associated trait information.

All interventions were implemented in September 2022, shortly before the main planting season. This design allows me to compare two kinds of information, cue-based inference and variety-specific trait information, and two channels for the latter, individual and community-level.

4.5 Measurement Approach and Empirical Validity

My measurement strategy is designed to address key identification challenges that plague the agricultural technology adoption literature. A key concern in prior work is the reliance on farmers' self-reported data to measure adoption and farmers' ability to identify variety traits. Such data suffer from several problems. First, self-reported data do not provide an objective measure of adoption when farmers misperceive the varieties they are growing (Wossen, Abdoulaye, et al., 2019; Wossen, Nguimkeu, et al., 2019). Second, even when farmers correctly identify the variety grown, they may misreport (Abay et al., 2021, 2023). Third, information treatments may alter reporting incentives leading to experimenter demand bias. To overcome these challenges, I use DNA fingerprinting to identify sweet potato varieties cultivated on sampled plots. This method allows for objective varietal identification and provides scientifically validated information on agronomic and nutritional traits. By linking DNA results to plot-level survey data and GPS-measured plot area,[13] I can precisely quantify adoption behavior and isolate treatment effects on the planting of specific varieties. This strategy captures realized planting behavior and reduces reliance on potentially biased self-reports.

[13]Plot area is another measure that has been shown to suffer from misreporting errors when farmers self-report (Carletto et al., 2017, 2015, 2013).

The timing of the endline survey, approximately six months after the intervention and shortly after the main sweet potato harvest, allows for identification of short-run effects and provides opportunities for robustness checks. For instance, if the treatments increase consumption of vitamin A-rich varieties, I should observe reductions in child diarrhea, an outcome responsive to micronutrient intake. In contrast, I should not expect effects on outcomes such as malaria or vision impairment, which are either unrelated to vitamin A or slow-moving, providing a clean placebo test. For this reason, I included in the survey not only diarrhea incidence but also fever, cough, difficulty breathing, malaria, and visual impairment. However, since health-related outcomes are self-reported, they may still be subject to experimenter demand bias. The presence of two treatment arms enables an indirect test of this concern. If responses were driven by experimenter demand bias, I would expect similar effects across arms regardless of content. Divergent treatment effects consistent with the theory of change and corroborated by objective adoption data would instead support a causal role for the informational content of the intervention.

Finally, I randomized the identity of the household respondent, who received the information on behalf of the household: a male adult, a female adult, or both jointly. Because every T2 recipient received the same DNA results, this randomization lets me test whether the planting response varies with how much the recipient values what the targeted traits deliver (ω_{ik} in the framework), holding what was learned fixed. I also use it to check whether recall of the session differs by respondent gender (Section 7.2, Table A2).

4.6 Attrition, Compliance, and Baseline Balance

Attrition. The endline survey achieved a high follow-up rate, successfully reinterviewing 2,109 of the 2,141 baseline households (98.5%). Table A3 examines whether attrition is systematically related to treatment assignment across multiple margins. Column (1) shows that neither treatment arm is significantly associated with attrition at endline, and the equality of coefficients across T1 and T2 cannot be rejected (p-value = 0.27). Column (2) indicates that treatment assignment does not predict whether households cultivated sweet potato in the second season, suggesting that subsequent outcomes conditional on sweet-potato cultivation are not differentially selected by treatment status. Column (3) examines attrition in the DNA-fingerprinting sample, which includes households that did not consent to the survey or leaf collection, attrited, or did not produce sweet potato in the 2023 second season, and again finds no systematic relationship with treatment assignment. Finally, Columns (4) and (5) test whether children sampled at baseline differentially attrited from the health and consumption modules, respectively; neither treatment arm significantly predicts

attrition in these modules. Taken together, these results suggest that differential attrition is unlikely to bias the estimated treatment effects across any of the main outcome domains.

Take-up. Intervention take-up rates were high across both treatment arms. In T1, 91% of targeted farmers attended the session, while in T2 the attendance rate was slightly higher at 93%. As shown in Column (6) of Table A3, this difference is not statistically significant (p-value = 0.15), indicating no differential compliance by treatment assignment. High attendance rates support the validity of the experimental design and reduce concerns related to differential exposure to the intervention.

Baseline Balance. Table A4 summarizes baseline characteristics by treatment arm. Most variables are well balanced across groups. Among household characteristics, T1 households are significantly larger than control households (p-value = 0.002) and marginally larger than T2 (p-value = 0.050). In child diet measures, control children were more likely to have consumed sweet potato in the previous seven days compared to both T1 (p-value = 0.010) and T2 (p-value = 0.030). There are also a few statistically significant differences in the baseline share of varieties: the share of yellow-fleshed varieties is higher in T2 than control (p-value = 0.016), and the share of AB resistant varieties differs between T1 and T2 (p-value = 0.030). Adoption of at least one orange-fleshed, yellow-fleshed, or disease-resistant variety is well balanced across treatment arms. The two imbalances in baseline varietal shares matter most for my main area outcomes, since a higher baseline share of yellow-fleshed or AB-resistant varieties could mechanically raise the area planted to them at endline. To address this and other baseline differences, all my specifications include controls selected through post-double-selection LASSO following Belloni et al. (2014), which flexibly adjusts for baseline characteristics, including baseline varietal shares. Given the limited number of clusters (108 villages) and the presence of accidental baseline imbalances, I additionally report randomization inference p-values for all main outcomes (adoption, consumption, and health). This approach accounts for the realized randomization and allows me to check that the estimated treatment effects are not artifacts of accidental baseline differences.

4.7 Empirical Strategy

I estimate intent-to-treat effects of the two treatments with the specification

$$y_{ihv} = \alpha + \beta_1 \, \mathrm{T1}_v + \beta_2 \, \mathrm{T2}_v + \mathbf{X}'_{hv}\gamma + \mu_s + \varepsilon_{ihv}, \tag{4}$$

where y_{ihv} is the outcome for unit i in household h and village v, $\mathrm{T1}_v$ and $\mathrm{T2}_v$ indicate the treatment arm of the village, with the control group omitted, μ_s are fixed effects for the randomization strata, and $\mathbf{X}_{hv}$ are the baseline controls selected by post-double-selection LASSO. The selection draws on a rich set of baseline variables: household demographics, agricultural characteristics, and detailed records of prior exposure to sweet-potato dissemination programs and health-related trainings over the decade preceding the intervention.[14] Standard errors are clustered at the village level, the unit of randomization.

The coefficient β_1 is the effect of the session on visual cues, and β_2 the effect of the session combined with variety-specific trait information, both relative to the control group. Their difference, $\beta_2 - \beta_1$, measures the additional effect of variety-specific trait information on top of the session; where a table reports a p-value for the equality of the two treatments, it tests this difference.

The unit of observation follows the outcome. Recall is measured for the respondent, and land allocation and production for the household. Identification is measured for each farmer–variety pair and as a household-level share. Consumption and health outcomes are measured for each child, and these specifications add child age fixed effects and a gender indicator.

To study heterogeneity, I add to (4) a baseline characteristic M_{hv} and its interactions with the two treatments,

$$y_{ihv} = \alpha + \beta_1 \,\mathrm{T1}_v + \beta_2 \,\mathrm{T2}_v + \delta \, M_{hv} + \theta_1 \,\mathrm{T1}_v \times M_{hv} + \theta_2 \,\mathrm{T2}_v \times M_{hv} + \mathbf{X}'_{hv}\gamma + \mu_s + \varepsilon_{ihv}, \quad (5)$$

so that β_2 is the effect of T2 for households with $M_{hv} = 0$ and $\beta_2 + \theta_2$ the effect for those with $M_{hv} = 1$. For spillovers, I estimate (4) with a single indicator for community-information households in T2 villages, on a sample in which the control group and T1 households form the comparison group.

Randomization inference p-values, reported for the main outcomes, are based on 1,000 permutations of the village-level assignment within strata.

5 Main Results

This section reports the empirical results, beginning with the main estimates of treatment effects on cue reading, correct identification of varietal traits, adoption, consumption patterns, and health outcomes.

The main outcomes are analyzed using the sample of farmers with DNA-fingerprinting

[14] Appendix D lists all variables used in the selection procedure.

data. Only in the next section do I use the full sample, to examine whether the session on visual cues, common to both treatments, taught farmers to read those cues. Section 5.7 then turns to spillover effects, drawing exclusively on the community-information households in T2 villages.

5.1 Reading the Cues

Both arms received the same session on visual cues, which taught farmers to read flesh color as a signal of nutritional content and plant symptoms as a signal of resistance to SPVD and AB; T2 additionally delivered DNA results on each farmer's own varieties. I first examine whether farmers retained this content, measured through unprompted recall, and whether they did so to a similar extent in both arms, so that the T2–T1 contrast reflects what variety-specific trait information adds on top of the session.

Table 3 reports effects on farmers' recall of the promoted information six months after the intervention. Relative to the control group, the share of respondents who recognize flesh color as a marker of nutritional value rises by 28 percent under T1 and 27 percent under T2 (Column 1), and the share who identify orange-fleshed varieties as the most nutritious rises by 65 percent under T1 and 78 percent under T2 (Column 2). Knowledge of disease symptoms and control practices for SPVD and AB increases by 0.5 to 1.1 standard deviations (Columns 3–6). All effects are statistically significant at the 1 percent level. The two arms differ only in knowledge of SPVD symptoms, where the effect is 1.07 standard deviations under T1 and 0.72 under T2 ($p = 0.04$). The DNA results in T2 therefore neither amplified nor crowded out the effect of the shared session.[15]

5.2 Correct Identification of Varietal Traits

I next examine whether the treatments changed how accurately farmers identify the traits of their varieties. The outcomes are "true positive" indicators, equal to one when a farmer reports that a variety has a given trait and DNA fingerprinting confirms it.

Table 4 reports results at the farmer–variety level for orange flesh, yellow flesh, SPVD resistance, and AB resistance. T1 has no significant effect on yellow flesh or on either resistance trait, and it reduces correct identification of orange-fleshed varieties ($p < 0.05$), though from a low base: 3 percent of control observations are true positives for orange flesh. Farmers rarely fail to recognize an orange-fleshed variety without the

[15] I compare the arms on recall because what T2 farmers say about their own varieties mixes what they learned from the session with what the DNA card told them, whereas recall of the session content reflects only the former.

intervention, with false negatives under 1 percent of control observations (Table 2). T2 increases correct identification of yellow-fleshed varieties by 68 percent ($p < 0.01$) and of AB-resistant varieties by 20 percent ($p < 0.10$), with no significant effect on orange flesh or SPVD resistance.[16]

Table 5 measures accuracy at the household level: the share of a household's varieties carrying a trait, according to the DNA results, that the farmer reports as having it. Relative to the control group, T2 farmers correctly identify 58 percent more of their yellow-fleshed varieties ($p < 0.05$) and 48 percent more of the varieties that combine yellow flesh and virus resistance ($p < 0.10$). T1 has no significant effect on any trait.

Across both tables, only T2 improves farmers' identification of the traits of their varieties. The gains are concentrated in yellow flesh, the trait control farmers most often misreport: 28 percent of control observations are false positives and 26 percent are false negatives (Table 4).

5.3 Varietal Adoption

I now examine whether the treatments changed the area farmers plant to varieties with the targeted traits.

Intensive Margin. Table 6 reports treatment effects on the area allocated to sweet potato varieties, grouped by the traits identified through DNA fingerprinting.[17] T1 has no statistically significant effect on the area planted to any category of varieties. T1 also did not improve trait identification (Section 5.2). I examine in Section 6 whether perceived trade-offs contribute to this result.

Relative to control means, T2 increases the area planted with yellow-fleshed varieties by 36 percent ($p = 0.06$), with SPVD-resistant varieties by 25 percent ($p = 0.09$), with AB-resistant varieties by 45 percent ($p = 0.02$), and with varieties that combine yellow flesh and virus resistance by 59 percent ($p = 0.02$). Randomization-inference p-values range from 0.01 to 0.10. Yellow flesh and AB resistance, the traits for which T2 improved identification, are among those with the largest area responses.

I do not observe a reduction in the area devoted to white- or purple-fleshed varieties, or to varieties not identified via DNA fingerprinting, about which the cards carried no information. The estimate for white-fleshed varieties is positive but imprecise (13

[16] These results may capture both improvement in trait recognition among varieties that were already grown and any shifts in the varieties grown.

[17] The outcome is constructed by multiplying GPS-measured plot size by the farmer's self-reported land share allocated to each variety. Varietal traits such as flesh color and virus resistance are assigned using DNA fingerprinting, not farmers' reports.

square meters, with a standard error of 24), so it does not rule out a moderate decline. These traditional varieties are often preferred for taste, an attribute the intervention did not target, and a variety is conserved only by planting it, so reductions may take more than one season. For the varieties farmers already hold, vegetative propagation allows planting material to be multiplied quickly, which likely enabled the expansions within one season.

The expansion is a net increase in sweet potato cultivation: T2 raises total sweet potato area by 18 percent ($p = 0.05$; Column 9), so farmers did not only reallocate land among the varieties they grow.

Extensive Margin. I do not detect treatment effects on whether farmers grow at least one variety with a given trait (Table A6), or on the number of such varieties grown (Table A7). The information arrived shortly before planting, which left farmers little time to obtain new varieties: without a formal seed market, a variety a farmer does not already grow can only be sourced from other farmers.

5.4 Production by Varietal Trait

Table 7 reports treatment effects on self-reported sweet potato quantities (consumed, harvested, still in the field, and total), by varietal traits identified through DNA fingerprinting. The quantities are reported by farmers from recall, while the varieties are classified using the DNA results.

T2 raises the quantities harvested and consumed of yellow-fleshed and virus-resistant varieties. Relative to the control mean, harvested quantities of yellow-fleshed varieties rise by 87 percent ($p < 0.05$) and consumed quantities by 67 percent ($p < 0.05$); the total rises by 51 percent ($p < 0.10$). For virus-resistant varieties, consumed quantities rise by 50 percent ($p < 0.05$) and harvested quantities by 50 percent ($p < 0.10$); the 33 percent increase in the total is not statistically significant. Totals are estimated less precisely because they include quantities still in the field at the time of the survey, which are noisy for a crop that is harvested gradually and often stored in the ground, and for which no effect is significant. These increases are in the trait groups for which T2 expanded area (Section 5.3).

Importantly, these increases in production of targeted varieties did not come at the expense of overall productivity. Total sweet potato yields show no statistically significant treatment effect (Table A8), providing little evidence for the substantial yield penalties farmers perceived (Figure 3). While I cannot rule out modest yield trade-offs, the data do not support the strong negative beliefs documented at baseline.

Under T1, the quantity consumed from yellow-fleshed varieties rises by 77 percent ($p < 0.10$), although T1 has no effect on the area planted to these varieties. This could

reflect changes in how farmers manage or harvest them, or reporting bias: farmers who retained the nutritional message may report higher consumption of yellow-fleshed varieties. The absence of T1 effects on area and on child health (Section 5.6) is more consistent with reporting bias.

5.5 Consumption Outcomes

This section examines sweet potato consumption among children who were under two and under five years of age at baseline, who face an elevated risk of vitamin A deficiency and therefore benefit the most from eating orange- and yellow-fleshed sweet potato (UNICEF, 2020; World Health Organization, 2009, 2011).

For children under two, Table 8 reports effects on consumption of orange- or yellow-fleshed sweet potato over two recall periods. T2 increases the probability that the child ate it in the previous seven days by 32 percent relative to the control mean ($p = 0.03$; RI $p = 0.09$). T2 also raises the probability of eating it at least once a week over the previous six months by 19 percent, and the number of times the child ate it in the previous seven days, but neither effect is statistically significant. T1 has no significant effect on any of the three outcomes.

For children under five, neither treatment has a significant effect (Table A9). One possible reason is that caregivers target older children less, because their diets are more varied and less directly shaped by messages emphasizing vitamin A intake.

To examine whether the intervention influenced broader dietary patterns for the youngest cohort, Table A10 reports effects on 24-hour food group consumption and the dietary diversity index. T1 again shows no systematic changes. T2 increases the likelihood of consuming fruit by 36 percent ($p = 0.01$; RI $p = 0.04$). Given the intervention's emphasis on orange color as a marker of vitamin A, this effect may reflect increased consumption of orange-colored fruits (e.g., mango, papaya), consistent with the nutrition message. Effects on other food categories are small and statistically insignificant, and the dietary diversity index shows no measurable change.

5.6 Health Outcomes

I examine whether the dietary changes induced by the interventions translated into improvements in child health. Vitamin A supports gut mucosal integrity and immune function, with clinical effects on diarrhea observable within weeks of increased intake (Mayo-Wilson et al., 2011; Stephensen, 2001). Broader immune effects, such as reduced respiratory infections, typically require sustained supplementation over longer periods (Imdad et al., 2022). With six months between the intervention and the endline, diarrhea is therefore the health outcome for which I expect measurable effects. I focus

on children under two at baseline, who are more vulnerable to vitamin A deficiency and for whom I find dietary changes (Section 5.5). I study three outcomes: (i) whether the child experienced diarrhea in the past two weeks, (ii) the duration of the most recent episode, and (iii) the incidence of diarrhea over the past six months. All regressions control for child age and gender.[18]

Table 9 reports results for children under two at baseline. T2 reduces the probability of diarrhea in the past two weeks by 52 percent relative to the control mean ($p = 0.03$; RI $p = 0.05$), and the duration of the most recent episode by 43 percent, or 1.3 days ($p = 0.002$; RI $p = 0.01$). The incidence of diarrhea over the past six months falls by 24 percent, but this estimate is not statistically significant ($p = 0.11$). T1 shortens the most recent episode by 30 percent, or 0.9 days ($p = 0.06$; RI $p = 0.11$), and has no significant effect on incidence over either recall period. The T2 coefficients are larger in magnitude than the T1 coefficients for all three outcomes, but the two arms are not statistically distinguishable from each other, likely because of the limited sample size.

Table A11 reports results for all children under five. I find no dietary changes for this group (Table A9), but both treatments reduce the duration of the most recent episode, by 20 percent under T1 ($p = 0.08$) and by 24 percent under T2 ($p = 0.06$), and T2 reduces six-month incidence by 22 percent ($p = 0.05$). These effects come from the children under two, for whom I find dietary changes and who account for most diarrhea morbidity in early childhood: among children aged 2–5, no effect is statistically significant (Table A12).

The 52 percent reduction in diarrhea is of the same order as the effects of earlier food-based vitamin A interventions in comparable settings. Jones and De Brauw (2015) documented a 42 percent reduction in diarrhea prevalence in their Mozambique trial promoting orange-fleshed sweet potato, with effects emerging within 6–9 months of the intervention. Biofortification trials in Uganda and Mozambique have documented increases in vitamin A intake and improved serum retinol concentrations (Hotz, Loechl, de Brauw, et al., 2012; Hotz, Loechl, Lubowa, et al., 2012), outcomes that lie along the causal pathway between dietary change and diarrhea reduction. The meta-analysis by Imdad et al. (2022), synthesizing 43 randomized trials of vitamin A supplementation, finds consistent reductions in diarrhea-related morbidity, particularly in populations with moderate-to-severe baseline deficiency.

Two further factors support the plausibility of these magnitudes in my setting. First, vitamin A deficiency has long been recognized as a public health concern among Ugandan children, though its measured prevalence is highly sensitive to survey

[18]The health sample includes up to four children per household, rather than the two-child restriction used in the dietary analysis. These additional children share the same household-level treatment assignment, mitigating concerns about compositional differences.

methodology: the 2011 UDHS reported that 33 percent of children aged 6–59 months were deficient, while the 2016 UDHS reported 9 percent, using the same biomarker, laboratory, and cut-off.[19] Dietary intake data are less sensitive to these issues and show shortfalls in both years: between one-third and two-fifths of children aged 6–23 months consumed no vitamin A-rich food in the 24 hours preceding the survey (39 percent in 2011, 34 percent in 2016) (Uganda Bureau of Statistics & ICF, 2018; Uganda Bureau of Statistics & ICF International Inc., 2012). This leaves scope for interventions that raise the vitamin A content of staple foods. Second, the six months between the intervention (September 2022) and the endline (February–March 2023) allowed treated households to plant, harvest, and consume orange- and yellow-fleshed varieties during the period in which child health was measured. The 32 percent increase in consumption of orange- or yellow-fleshed sweet potato among children under two (Table 8) is the proximate mechanism linking the intervention to health outcomes.

5.7 Spillovers to Community-Information Households

Delivering variety-specific trait information through DNA fingerprinting to every household is logistically demanding and costly. This raises a policy question central to scalability: can information delivered to only a subset of farmers generate spillovers to the rest of their community? Within T2 villages, I shared cards listing the individual-information households, their reported and DNA-verified variety names, and the associated traits with the community-information households, farmers in the same villages who were not themselves sampled. If indirect exposure alone shifts behavior, supplying community-level information from a subsample would be cheaper than sampling every household and delivering individual results.

I cannot measure adoption by DNA-identified trait for community-information households, since this requires their own DNA results. I therefore examine downstream outcomes, sweet potato consumption and child health, comparing community-information households to a comparison group pooling the pure control group and the T1 group, which received the cues session but no DNA results. This rests on the assumption that indirect exposure operates through the same channels as among

[19] Both surveys assessed vitamin A status via retinol-binding protein enzyme immunoassay on dried blood spots (cut-off 0.825 μmol/l, equivalent to 0.7 μmol/l serum retinol), adjusted for inflammation following Thurnham, McCabe, Northrop-Clewes, and Nestel (2003). The 2011 estimate is 38% unadjusted and 33% adjusted (Uganda Bureau of Statistics & ICF International Inc., 2013); the 2016 figures are 15 and 9% (Uganda Bureau of Statistics & ICF, 2018). The 2011 report attributes its elevated estimate to delayed sample transport and slowed drying during an unusually heavy rainy season, which may have degraded the blood spots. Baingana, Matovu-Kasozi, and Garrett (2013) similarly show that failing to adjust for the acute-phase response overstated deficiency in the 2006 UDHS by roughly 25%.

individual-information households: changes in what households plant and feed their children.

For children under two, Table 10 shows that the probability of consuming orange or yellow sweet potato in the past seven days increases by 37 percent relative to the comparison-group mean ($p < 0.10$). Effects on consumption frequency and over the longer recall period are positive but not statistically significant. Disaggregating by terciles of the baseline village-level share of orange-fleshed or virus-resistant varieties (Table A13), the point estimates are larger in the second and third terciles, though none is statistically significant.[20]

Turning to health, Table 11 reports that the probability of diarrhea in the past two weeks falls by 61 percent among children under two in community-information households ($p < 0.05$). Effects on episode duration and over the longer recall period are negative but imprecisely estimated.[21] By tercile, the reduction in two-week incidence is statistically significant only in villages with the highest baseline share of these varieties, although I cannot reject that the effect is the same across terciles (Table A15).

Taken together, these results provide suggestive evidence that community-level dissemination of DNA-based varietal information generates spillover benefits for community-information households, with effects on consumption and on diarrhea that are similar in size to those found among individual-information households.

6 Perceived Trade-offs in Varietal Traits

Section 5.3 showed that T1 did not change what farmers plant. One possible explanation is that a targeted trait has little value to a farmer who believes it comes at the expense of something else, and Section 2.2 showed that farmers do hold such beliefs. I provide two pieces of evidence. First, I examine whether T1 changed the area farmers plant to the varieties they themselves report as carrying the targeted traits. Second, I examine whether the response to T2 is larger where the information shows a variety that combines the two targeted traits (a "win-win" variety).

[20] Terciles partition villages into thirds by their baseline share of DNA-verified orange-fleshed or virus-resistant varieties, from the lowest third (first tercile) to the highest (third tercile). Because villages were stratified on this share, together with region, at randomization, the terciles reflect pre-assigned variation in the baseline prevalence of advantageous varieties rather than a post-hoc partition, and they proxy for the potential intensity of the treatment, since the community card in higher-share villages covers a larger set of locally relevant varieties.

[21] No significant effects appear among children under five (Table A14), consistent with the results for individual-information households.

6.1 Reading the Cues Alone Is Insufficient in the Presence of Trade-offs

T1 taught farmers to read the cues but did not improve their identification of which varieties carry which traits (Table 4). Area by DNA-identified trait cannot show a response to T1 if farmers apply the cues to the wrong varieties, so I use area by self-reported trait. If learning to read the cues were enough to change planting, T1 farmers would expand the varieties they themselves report as carrying the targeted traits. Table 12 shows that they did not: T1 has no statistically significant effect on the area planted to varieties that farmers report as orange-fleshed, yellow-fleshed, SPVD-resistant, or AB-resistant, although the estimates are imprecise. Because this outcome follows the traits as farmers report them, a misreading of the cues cannot explain the result. The result is consistent with farmers seeing little gain in a variety they regard as nutritious or resistant when they believe that the trait comes at the expense of other traits. The next subsection provides more direct evidence on trade-offs.

6.2 Correcting Trade-off Misperceptions Unlocks Adoption

I next examine whether the planting response to T2 is larger where the information contradicts perceived trade-offs. Using baseline data, I construct an indicator for whether a household (Table 13) or a village (Table 14) had at least one win-win variety before treatment, and interact it with treatment assignment.

In Table 13, the outcome is the area allocated to varieties grouped by trait. The main coefficients give the effects for households without a win-win variety, and the interaction (T2 $\times$ Win-win) gives the additional effect for households that had at least one DNA-verified win-win variety on their farm at baseline. For households without such a variety, the T2 coefficients are not statistically significant in any column. For households with one, T2 raises the area planted to yellow-fleshed varieties by an additional 85 square meters ($p < 0.05$), to SPVD-resistant varieties by an additional 167 ($p < 0.01$), and to AB-resistant varieties by an additional 141 ($p < 0.05$).

There are two ways to read this pattern. The card corrects the farmer's beliefs about both traits of such a variety at once. Seeing that one of her own varieties combines the two traits may also weaken her belief that desirable traits trade off more generally. The second reading has an implication of its own: her other varieties that carry a targeted trait should gain as well. In the same table, T2 also raises the area under white-fleshed varieties among these households, by 96 square meters relative to T2 households without a win-win variety ($p < 0.05$), and the large majority of DNA-matched varieties are resistant to SPVD (Table 2). This is consistent with the

second reading, with the caution that applies to the rest of this analysis. T2 also improves farmers' ability to identify the varieties that combine yellow flesh and virus resistance (Table 5), which is the information needed to revise a belief that these traits trade off. One caveat is that a household with a win-win variety also has an attractive variety to expand once T2 reveals which one it is. The interaction therefore does not separate the correction of trade-off beliefs from the availability of a variety with both targeted traits.

Looking within win-win households helps to address this. This analysis, like the win-win interaction above, was not pre-specified and should be read as exploratory. Table A16 restricts the sample to the 224 DNA-sampled households that had at least one win-win variety at baseline, and interacts treatment with the share of those varieties that the farmer did not report as combining the two traits (false negatives). Among farmers who recognized all of their win-win varieties, the T2 effects are positive but not statistically significant. Among farmers who recognized none of them, T2 raises the area planted to yellow-fleshed varieties by 66 percent of the control mean in this sample (128 square meters, $p = 0.02$), to SPVD-resistant varieties by 62 percent (177 square meters, $p = 0.04$), to AB-resistant varieties by 101 percent (211 square meters, $p < 0.01$), and total sweet potato area by 48 percent (296 square meters, $p = 0.01$). The difference between the two groups is estimated imprecisely, and the interaction is significant at the 10 percent level only for yellow flesh. I therefore read the response as concentrated among, but not confined to, farmers for whom the information revealed something new: a farmer who already knew that she held such a variety had less to learn from the card. T1 shows no positive response in either group.[22]

Table 14 repeats the specification with the win-win indicator defined at the village level, equal to one if any farmer in the village grew such a variety before treatment. I expect weaker effects here, because a variety grown by a neighbor and listed on the community card is weaker evidence against the trade-off than a variety on the farmer's own card, and because a given household is less likely to grow or recognize it. The interactions are smaller than at the household level and statistically significant only for yellow-fleshed varieties, at the 10 percent level, but they have the same sign for the varieties with the targeted traits.

[22] Among farmers who recognized all of their win-win varieties, the T1 estimates are negative (-152 square meters for yellow-fleshed varieties, $p < 0.05$).

7 Heterogeneous Treatment Effects

7.1 Heterogeneous Effects by Baseline Information Intensity

Table 15 examines how treatment effects on varietal planting vary across terciles of the baseline village-level share of orange-fleshed or virus-resistant varieties. Because this variable was part of the stratification in the experimental design, the terciles capture systematic differences in the *potential intensity* of the information treatment. In villages with a higher baseline share of nutritionally advantageous or disease-resistant varieties, the DNA results covered a larger number of locally grown varieties, making the intervention effectively more informative and actionable. Conversely, in villages where such varieties were rare at baseline, the actionable information delivered to farmers was necessarily more limited.

The results are consistent with this interpretation. T2 generates very large and statistically significant increases in area planted to yellow-fleshed and virus-resistant varieties only in the third tercile, where the baseline prevalence of improved varieties was highest. In these villages, T2 leads to substantial expansion in land area devoted to yellow-flesh, SPVD-resistant, AB-resistant, and combined yellow-and-virus-resistant varieties, with effect sizes exceeding 100 square meters. These strong responses suggest that when the informational treatment is more intensive, because more varieties can be accurately identified and reported, farmers respond by expanding the area under the varieties identified as carrying the targeted traits.

In contrast, T2 effects are small and often negative in the first tercile, where villages had very low baseline shares of orange or virus-resistant varieties. In these settings, the intervention delivered relatively little actionable varietal information, which may explain the absence of meaningful behavioral change. The second tercile shows no systematic pattern, consistent with an intermediate level of information intensity.

T1 effects are weaker and show no consistent gradient across terciles. This is what I would expect: the cues session, without variety-specific information, does not become more informative where such varieties are more common.

The joint tests reinforce these conclusions: treatment effects differ significantly across terciles for T2, especially between the highest and the lower terciles, whereas T1 exhibits no comparable pattern. This suggests that the effective dose of informational treatment depends critically on the baseline varietal landscape. Where many relevant varieties are present, the intervention delivers richer information; where few are present, the informational content is thin and less actionable, at least in the short run.

7.2 Heterogeneous Effects by Trait Valuation: Respondent Gender

The framework implies that, for a given correction of beliefs, the planting response to T2 scales with the value the recipient places on what the targeted traits deliver, ω_{ik} (Prediction 4). The respondent randomization described in Section 4.5 provides a test. In Uganda, women are both the primary sweet potato cultivators and the main caregivers responsible for child feeding, so nutritional and disease-resistance traits are plausibly more valuable to them than to men. Because every T2 recipient received the same DNA results, differences in the planting response by recipient gender reflect differences in valuation rather than in what was learned.

Table 16 presents treatment effects disaggregated by respondent gender.[23] T1 generated no significant adoption responses for either men or women. For T2, the main effects, which capture men's responses, remain small and insignificant across all trait categories. In contrast, women receiving T2 increased the area under yellow-fleshed varieties by 47 square meters ($p = 0.02$), SPVD-resistant varieties by 65 square meters ($p = 0.05$), AB-resistant varieties by 73 square meters ($p = 0.01$), and varieties combining both traits by 48 square meters ($p = 0.01$), increases of 38%, 30%, 45%, and 59% relative to the control mean among male-respondent households. Total sweet potato area also expanded by 99 square meters ($p = 0.06$), indicating that women expanded overall cultivation rather than only reallocating existing land.

The gender gap in planting responses does not reflect a gap in learning. Recall of the session is similar for men and women in both arms (Table A2): men's recall rises significantly on every outcome, and where women recall slightly more, as for the color–nutrition rule under T2, the difference is too small to account for the planting gap, which is just as large for the resistance traits where recall does not differ. With beliefs corrected equally under T2, the difference in planting reflects how much the recipient values the traits. A remaining possibility is that respondent identity also affects whether the information reaches the person who decides what to plant, or how freely that person can move land between varieties; equal recall across genders does not rule these out, and I read the heterogeneity as consistent with, rather than proof of, differences in valuation.

[23] Female respondent equals one if the assigned respondent was a woman, either alone or jointly with a man, and zero if a man alone.

8 Robustness Tests

In this section, I assess the robustness of the main consumption and health results presented above. I focus on evaluating the plausibility of alternative explanations for the observed increase in sweet potato consumption and health improvements among young children and consider whether these patterns could be attributed to differential changes in reporting behavior, previous exposure to trainings or other unobserved factors.

8.1 Ruling Out Misclassification Bias in Orange/Yellow-Fleshed Sweet Potato Consumption

The observed increase in orange/yellow-fleshed sweet potato consumption is consistent with the increases observed in both the area planted and the quantities harvested, as objectively verified through DNA fingerprinting. Nevertheless, because consumption data are self-reported, it is important to assess whether treatment-induced differences in farmers' varietal identification ability could have biased these results. Specifically, if farmers in the treatment group are better at correctly identifying yellow-fleshed varieties at endline than control farmers, this could lead to higher reported consumption without corresponding behavioral changes.

I address this concern with a bounding analysis, following Lewbel (2007) and Kreider and Pepper (2007), that asks how large the treatment–control difference in misclassification rates would have to be to fully explain the observed consumption effect, and compares it with the difference actually observed at endline, where DNA fingerprinting validates farmers' self-reported varietal identification. The analysis focuses on T2, the arm with significant effects on both consumption and varietal identification. Appendix E presents the framework, the scenarios, and the full results (Table A18).

Under the scenario most favorable to the misclassification explanation, the observed difference in false negative rates (5.3 pp) accounts for at most 36% of the 14.6 pp treatment effect; in the scenarios where false positives would have to do the work, the observed difference (0.07 pp) is more than 200 times smaller than required. Because households consume mainly from their own production (93%), production-side misclassification is a relevant proxy for consumption misreporting, and the absence of extensive-margin effects limits the scope for false positives. Treatment-induced differences in varietal recognition may therefore explain part of the increase in reported consumption, but not the treatment effect itself.

8.2 Placebo Health Outcomes

To assess the internal validity of my main health findings and rule out alternative explanations, such as experimenter demand effects or spurious shifts in reporting, I examine a set of non-targeted health outcomes included in the survey for this purpose. Although vitamin A supports immune function and may reduce overall morbidity, such effects are unlikely to manifest uniformly across health domains within the short time frame of my study. I therefore report treatment effects on outcomes that are not expected to respond to dietary vitamin A intake over a six-month horizon, including respiratory symptoms (difficulty breathing and cough), fever, malaria, and visual impairment.

Notably, while the intervention highlighted the general health benefits of vitamin A, specifically its role in immune function and maintaining night vision, it did not mention diarrhea or gastrointestinal illness. If respondents were adjusting their reporting based on these highlighted domains, I would expect improvements across all referenced symptoms. Instead, the pattern of results points to specificity consistent with biological mechanisms rather than generalized reporting bias.

Table 17 presents treatment effects on respiratory and febrile symptoms for children under two, the subgroup in which dietary change and diarrhea reductions were most pronounced. Most estimates are small and statistically insignificant. The only exception is a modest and marginally significant decline in fever incidence under T2.

Table 18 reports treatment effects on night-vision problems for children under 16. I examine two outcomes: whether the child has difficulty seeing at night, and whether they both report difficulty and wear corrective glasses. These outcomes correspond to one of the few health benefits mentioned in the intervention script, yet I find no evidence of effects on either measure.

These patterns align with findings from the clinical and epidemiological literature. For respiratory outcomes, vitamin A supplementation trials have produced mixed results: a meta-analysis by Grotto et al. (2003) finds no consistent reductions in acute respiratory infections, and some studies report increased risk among children without deficiency (Stephensen, 2001). Similarly, although vitamin A modulates immune responses, including cytokine signaling and T-cell function, it does not target *Plasmodium* parasites or alter vector exposure. Large randomized trials in malaria-endemic settings find no effect on malaria incidence or parasite density (Binka et al., 1995; Shankar et al., 1999).

The Cochrane review cited above (Imdad et al., 2022) likewise finds no consistent effects of supplementation on malaria, respiratory infections, or measles-related morbidity, while confirming reductions in diarrhea. Finally, because fever is a non-specific symptom arising from many pathogens, including enteric infections, the modest decline

observed under T2 may plausibly reflect the documented reduction in diarrhea.

Taken together, these results strengthen the credibility of my main health findings. The absence of treatment effects on non-targeted outcomes, including symptoms explicitly referenced in the intervention, supports the interpretation that improvements in child health were specific to vitamin A-mediated protection against gastrointestinal infections, rather than artifacts of reporting bias or generalized behavioral responses.

8.3 Do Prior Trainings Explain the Health Effects?

An important question for interpreting the health impacts of the varietal identification intervention is whether the observed reductions in diarrhea among young children reflect the activation of prior knowledge that households had acquired before the study. If T2 merely reinforced or reactivated earlier exposure to nutrition or health messages, rather than providing substantively new information, then households with prior training should exhibit systematically larger improvements in health outcomes. To assess this possibility, I examine heterogeneous effects by whether the household reported attending any nutrition or health training in the five years preceding baseline.

Table 19 presents the interaction between treatment assignment and baseline exposure to nutrition training for diarrhea incidence and duration among children under age two, the group for whom I documented the strongest dietary and health effects. Across all three outcomes, I find no evidence that prior training amplifies the impact of either T1 or T2. The interaction terms are small, imprecisely estimated, and consistently statistically insignificant. For T1, the combined effect of the main treatment coefficient and the interaction term is not statistically different from zero across all outcomes. For T2, which generated substantial reductions in diarrhea, the interaction patterns are similarly flat: households with prior training do not experience larger effects than those without it, and the joint tests confirm that the T2 effect remains significant even among households without any training exposure. These results indicate that pre-existing nutrition or health knowledge is not driving the observed improvements.

9 Conclusion

This paper shows that variety misidentification is a critical but under-recognized constraint on agricultural technology adoption. Using a randomized controlled trial with sweet potato farmers in Uganda, I document that farmers frequently hold inaccurate beliefs about the traits of the varieties they cultivate, and that this inaccuracy has behavioral consequences. Teaching farmers to read the visual cues to

nutritional and disease-resistance traits improves what they know but does not shift planting; only variety-specific trait information on farmers' own varieties, which corrects their beliefs, changes adoption. These corrected choices increase the area planted to varieties combining nutritional and resilience traits, raise harvested quantities of yellow-fleshed varieties, and improve diet and diarrhea outcomes among the youngest children, the group for whom the dietary changes are concentrated.

These findings carry three implications, each of which should be read against the limits of a single trial and of a design without a DNA-only arm, which identifies the effect of variety-specific trait information on top of the cues session rather than on its own. First, accurate varietal identification may be a high-value input into agricultural policy: where farmers misperceive what they grow, their varietal choices are systematically distorted, eroding the returns to investments in breeding, extension, and seed systems. Embedding low-cost varietal diagnostics into extension services could unlock latent demand for existing high-performing varieties. Second, by revealing which traits farmers value once misperceptions are corrected, varietal identification can help orient breeding toward attribute combinations that match farmer preferences and local constraints. Third, the spillovers I document suggest that information delivered to a subset of farmers may generate community-level benefits. The cost accounting in Appendix F is consistent with this: the diagnostic, at $21 to $26 per sampled farmer, is the only component that scales with the number of households tested, and the binding margin for low-cost delivery is the coverage of the reference variety library rather than the share of farmers tested.

Beyond sweet potato, these results suggest that variety choice is a foundational decision in agricultural production systems. I deliberately study a low-input, vegetatively propagated crop in order to isolate trait recognition from decisions about complementary inputs and management, but in input-responsive crops the two are linked. In Ethiopian maize, farmers who correctly identify their improved variety substantially increase input use (urea, NPS, and labor) and yields, though without detectable changes in other practices or in profits (Mallia, 2022). This complementary, quasi-experimental evidence suggests that the downstream consequences of misidentification may be *larger* in input-intensive systems, where varietal traits interact with fertilizer, pest-management, and input-intensity decisions, an implication the present design cannot test directly. More broadly, the frictions identified here, misidentification of a technology, mistaken beliefs about how its attributes trade off, and the gap between reading a cue and accurate identification, are not specific to agriculture. They may arise wherever users adopt multi-attribute technologies whose traits are imperfectly observable and must be inferred from noisy signals, as in health products, climate-adaptation technologies, or tools used by small enterprises. In such settings,

generic information campaigns may do little unless they correct the underlying misperceptions, while low-cost variety-specific information could help align beliefs with actual product characteristics. Establishing the scalability of such diagnostics, the persistence of belief updating, and the breadth of these mechanisms across crops and domains is a promising direction for future work.

References

Abay, K. A., Bevis, L. E. M., & Barrett, C. B. (2021). Measurement error mechanisms matter: Agricultural intensification with farmer misperceptions and misreporting. *American Journal of Agricultural Economics*, *103*(2), 498-522. Retrieved from https://onlinelibrary.wiley.com/doi/10.1111/ajae.12173 doi: 10.1111/ajae.12173

Abay, K. A., Wossen, T., Abate, G. T., Stevenson, J. R., Michelson, H., & Barrett, C. B. (2023). Inferential and behavioral implications of measurement error in agricultural data. *Annual Review of Resource Economics*, *15*, 63-83. Retrieved from https://www.annualreviews.org/content/journals/10.1146/annurev-resource-101422-090049 doi: 10.1146/annurev-resource-101422-090049

Baingana, R., Matovu-Kasozi, D., & Garrett, D. (2013). The importance of controlling for the acute-phase response in the population-based assessment of vitamin a status: a study in children of pre-school age in Uganda. *Public Health Nutrition*, *16*(9), 1540–1547. doi: 10.1017/S1368980012004351

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

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, *81*(2), 608–650.

Binka, F. N., Ross, D. A., Morris, S. S., Kirkwood, B. R., Arthur, P., Dollimore, N., ... Smith, P. G. (1995). Vitamin A supplementation and childhood malaria in northern Ghana. *The American Journal of Clinical Nutrition*, *61*(4), 853–859. doi: 10.1093/ajcn/61.4.853

Bohr, N., Deisemann, T., Gollin, D., Kosmowski, F., & Lybbert, T. J. (2024). The seeds of misallocation: Fertilizer use and maize varietal misidentification in ethiopia. *Journal of Development Economics*, *171*, 103349. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304387824000981 doi: https://doi.org/10.1016/j.jdeveco.2024.103349

Bold, T., Kaizzi, K. C., Svensson, J., & Yanagizawa-Drott, D. (2017). Lemon technologies and adoption: Measurement, theory and evidence from agricultural markets in uganda. *The Quarterly Journal of Economics*, *132*(3), 1055–1100. doi: 10.1093/qje/qjx009

Bouis, H. E., & Saltzman, A. (2017). Improving nutrition through biofortification: A review of evidence from HarvestPlus, 2003 through 2016. *Global Food Security*, *12*, 49–58. Retrieved 2025-06-05, from https://linkinghub.elsevier.com/

retrieve/pii/S2211912417300068 doi: 10.1016/j.gfs.2017.01.009

Carletto, C., Gourlay, S., Murray, S., & Zezza, A. (2017, Oct.). Cheaper, faster, and more than good enough: Is gps the new gold standard in land area measurement? *Survey Research Methods*, *11*(3), 235–265. Retrieved from https://ojs.ub.uni-konstanz.de/srm/article/view/6791 doi: 10.18148/srm/2017.v11i3.6791

Carletto, C., Gourlay, S., & Winters, P. (2015, 05). From guesstimates to gpstimates: Land area measurement and implications for agricultural analysis. *Journal of African Economies*, *24*(5), 593-628. Retrieved from https://doi.org/10.1093/jae/ejv011 doi: 10.1093/jae/ejv011

Carletto, C., Savastano, S., & Zezza, A. (2013). Fact or artifact: The impact of measurement errors on the farm size–productivity relationship. *Journal of Development Economics*, *103*, 254-261. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304387813000345 doi: 10.1016/j.jdeveco.2013.03.004

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

Coomes, O. T., McGuire, S. J., Garine, E., Caillon, S., McKey, D., Demeulenaere, E., ... Wencélius, J. (2015). Farmer seed networks make a limited contribution to agriculture? four common misconceptions. *Food Policy*, *56*, 41-50. Retrieved from https://www.sciencedirect.com/science/article/pii/S030691921500086X doi: https://doi.org/10.1016/j.foodpol.2015.07.008

Corral, C., Giné, X., Mahajan, A., & Seira, E. (2020). *Autonomy and Specificity.* (Working paper)

de Brauw, A., Eozenou, P., Gilligan, D. O., Hotz, C., Kumar, N., & Meenakshi, J. V. (2018). Biofortification, crop adoption and health information: Impact pathways in mozambique and uganda. *American Journal of Agricultural Economics*, *100*(3), 906–930. doi: 10.1093/ajae/aay005

DellaVigna, S. (2009, June). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, *47*(2), 315–72. Retrieved from https://www.aeaweb.org/articles?id=10.1257/jel.47.2.315 doi: 10.1257/jel.47.2.315

Emerick, K., de Janvry, A., Sadoulet, E., & Dar, M. H. (2016). Technological innovations, downside risk, and the modernization of agriculture. *American Economic Review*, *106*(6), 1537–1561.

Evenson, R. E., & Gollin, D. (2003). Assessing the impact of the green revolution, 1960 to 2000. *Science*, *300*(5620), 758–762. doi: 10.1126/science.1078710

Floro, V. O., Labarta, R. A., Becerra López-Lavalle, L. A., Martinez, J. M., & Ovalle, T. M. (2018). Household determinants of the adoption of improved cassava varieties using DNA fingerprinting to identify varieties in farmer fields: A case

study in Colombia. *Journal of Agricultural Economics*, *69*(2), 518–536. doi: 10.1111/1477-9552.12247

Foster, A. D., & Rosenzweig, M. R. (1995). Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *Journal of Political Economy*, *103*(6), 1176–1209. Retrieved 2024-01-26, from https://www.journals.uchicago.edu/doi/10.1086/601447 doi: 10.1086/601447

Foster, A. D., & Rosenzweig, M. R. (2010). Microeconomics of Technology Adoption. *Annual Review of Economics*, *2*(1), 395–424. Retrieved 2025-08-10, from https://www.annualreviews.org/doi/10.1146/annurev.economics.102308.124433 doi: 10.1146/annurev.economics.102308.124433

Gollin, D., Hansen, C. W., & Wingender, A. M. (2021). Two blades of grass: The impact of the green revolution. *Journal of Political Economy*, *129*(8), 2344–2384. doi: 10.1086/714444

Grotto, I., Mimouni, M., Gdalevich, M., & Mimouni, D. (2003). Vitamin A supplementation and childhood morbidity from diarrhea and respiratory infections: A meta-analysis. *The Journal of Pediatrics*, *142*(3), 297–304.

Hanna, R., Mullainathan, S., & Schwartzstein, J. (2014). Learning Through Noticing: Theory and Evidence from a Field Experiment*. *The Quarterly Journal of Economics*, *129*(3), 1311–1353. Retrieved 2024-02-05, from https://academic.oup.com/qje/article/129/3/1311/1817927 doi: 10.1093/qje/qju015

HarvestPlus. (2012). *Disseminating Orange-Fleshed Sweet Potato: Findings from a HarvestPlus Project in Mozambique and Uganda.* (Tech. Rep.). Washington, DC: HarvestPlus.

Hotz, C., Loechl, C., de Brauw, A., Eozenou, P., Gilligan, D., Moursi, M., ... Meenakshi, J. V. (2012). A large-scale intervention to introduce orange sweet potato in rural Mozambique increases vitamin A intakes among children and women. *British Journal of Nutrition*, *108*(1), 163–176. doi: 10.1017/S0007114511005174

Hotz, C., Loechl, C., Lubowa, A., Tumwine, J. K., Ndeezi, G., Nandutu Masawi, A., ... Gilligan, D. O. (2012). Introduction of β-carotene-rich orange sweet potato in rural Uganda resulted in increased vitamin A intakes among children and women and improved vitamin A status among children. *The Journal of Nutrition*, *142*(10), 1871–1880. doi: 10.3945/jn.111.151829

Ilukor, J., Letaa, E., Khanal, A., Barros, J., Taye, L., Gimode, D., ... Macours, K. (2025). *SPIA Uganda Report 2025: Agricultural Diversity Under Stress.* (Tech. Rep.). Rome: CGIAR Independent Advisory and Evaluation Service (IAES).

Imdad, A., Mayo-Wilson, E., Haykal, M. R., Regan, A., Sidhu, J., Smith, A., & Bhutta, Z. A. (2022). Vitamin A supplementation for preventing morbidity

and mortality in children from six months to five years of age. *Cochrane Database of Systematic Reviews*, *2025*(4). Retrieved 2025-05-11, from http://doi.wiley.com/10.1002/14651858.CD008524.pub4 doi: 10.1002/14651858.CD008524.pub4

Jack, B. (2011). *Market inefficiencies and the adoption of agricultural technologies in developing countries.* (White paper prepared for the Agricultural Technology Adoption Initiative)

Jaleta, M., Tesfaye, K., Alemu, D., Diro, M., Yirga, C., Erenstein, O., & Abate, T. (2020). Misidentification by farmers of the crop varieties they grow: Lessons from dna fingerprinting of wheat in ethiopia. *PLoS ONE*, *15*(1), e0227044.

Jones, K. M., & De Brauw, A. (2015). Using Agriculture to Improve Child Health: Promoting Orange Sweet Potatoes Reduces Diarrhea. *World Development*, *74*, 15–24. Retrieved 2024-01-26, from https://linkinghub.elsevier.com/retrieve/pii/S0305750X15000911 doi: 10.1016/j.worlddev.2015.04.007

Kosmowski, F., Alemu, S., Mallia, P., Stevenson, J., & Macours, K. (2020). *Shining a Brighter Light: Comprehensive Evidence on Adoption* (Tech. Rep.). Rome: Standing Panel on Impact Assessment (SPIA).

Kreider, B., & Pepper, J. V. (2007). Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors. *Journal of the American Statistical Association*, *102*(478), 432–441. Retrieved 2025-05-16, from http://www.tandfonline.com/doi/abs/10.1198/016214506000000997 doi: 10.1198/016214506000000997

Laajaj, R., & Macours, K. (2026). The complexity of multidimensional learning in agriculture. *Econometrica*, *94*(2), 465-503. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA22974 doi: 10.3982/ECTA22974

Lewbel, A. (2007). Estimation of Average Treatment Effects with Misclassification. *Econometrica*, *75*(2), 537–551. Retrieved 2025-05-16, from http://doi.wiley.com/10.1111/j.1468-0262.2006.00756.x doi: 10.1111/j.1468-0262.2006.00756.x

Louwaars, N. P., & de Boef, W. S. (2012). Integrated seed sector development in africa: A conceptual framework for creating coherence between practices, programs, and policies. *Journal of Crop Improvement*, *26*(1), 39–59. doi: 10.1080/15427528.2011.611277

Low, J. W., Arimond, M., Osman, N., Cunguara, B., Zano, F., & Tschirley, D. (2007). A Food-Based Approach Introducing Orange-Fleshed Sweet Potatoes Increased Vitamin A Intake and Serum Retinol Concentrations in Young Children in Rural Mozambique ,3. *The Journal of Nutrition*, *137*(5), 1320–1327. Retrieved 2025-05-17, from https://linkinghub.elsevier.com/retrieve/pii/

S0022316622092288 doi: 10.1093/jn/137.5.1320

Macours, K. (2019). Farmers' Demand and the Traits and Diffusion of Agricultural Innovations in Developing Countries. *Annual Review of Resource Economics*, *11*(1), 483–499. Retrieved 2024-01-26, from https://www.annualreviews.org/doi/10.1146/annurev-resource-100518-094045 doi: 10.1146/annurev-resource-100518-094045

Macours, K., Mallia, P., & Okello, J. J. (n.d.). Large-Scale Roll-Out and Impacts of Biofortified Crop Varieties in Uganda. *Mimeo*.

Magruder, J. R. (2018). An assessment of experimental evidence on agricultural technology adoption in developing countries [Journal Article]. *Annual Review of Resource Economics*, *10*(Volume 10, 2018), 299-316. Retrieved from https://www.annualreviews.org/content/journals/10.1146/annurev-resource-100517-023202 doi: https://doi.org/10.1146/annurev-resource-100517-023202

Mallia, P. (2022). *You reap what (you think) you sow? evidence on farmers' behavioral adjustments in the case of correct crop varietal identification.* Retrieved from https://hal.science/hal-03597332v2 (Working paper, HAL-03597332v2)

Maredia, M. K., Reyes, B. A., Manu-Aduening, J., Dankyi, A., Hamazakaza, P., Muimui, K., ... Raatz, B. (2016). *Testing alternative methods of varietal identification using DNA fingerprinting: Results of pilot studies in Ghana and Zambia* (MSU International Development Working Paper No. 149). East Lansing: Michigan State University, Department of Agricultural Economics.

Mayo-Wilson, E., Imdad, A., Herzer, K., Yakoob, M. Y., & Bhutta, Z. A. (2011). Vitamin A supplements for preventing mortality, illness, and blindness in children aged under 5: Systematic review and meta-analysis. *BMJ*, *343*, d5094-d5094. Retrieved 2024-01-26, from https://www.bmj.com/lookup/doi/10.1136/bmj.d5094 doi: 10.1136/bmj.d5094

McGuire, S., & Sperling, L. (2016). Seed systems smallholder farmers use. *Food Security*, *8*(1), 179–195. Retrieved from https://doi.org/10.1007/s12571-015-0528-8 doi: 10.1007/s12571-015-0528-8

Mwanga, R. O. M., Andrade, M. I., Carey, E. E., Low, J. W., Yencho, G. C., & Grüneberg, W. J. (2017). Sweetpotato (Ipomoea batatas L.). In H. Campos & P. D. S. Caligari (Eds.), *Genetic improvement of tropical crops* (pp. 181–218). Cham, Switzerland: Springer International Publishing AG.

Mwanga, R. O. M., Mayanja, S., Swanckaert, J., Nakitto, M., zum Felde, T., Grüneberg, W., ... Yencho, G. C. (2021). Development of a food product profile for boiled and steamed sweetpotato in Uganda for effective breeding. *International Journal of Food Science & Technology*, *56*(2), 648–659. doi: 10.1111/ijfs.14721

Pingali, P. L. (2012). Green revolution: Impacts, limits, and the path ahead. *Proceedings of the National Academy of Sciences*, *109*(31), 12302–12308. doi: 10.1073/pnas.0912953109

Poets, A., Silverstein, K., Pardey, P., Hearne, S., & Stevenson, J. (2020). *DNA fingerprinting for crop varietal identification: Fit-for-purpose protocols, their costs and analytical implications* (Tech. Rep.). Rome, Italy: Standing Panel on Impact Assessment (SPIA). Retrieved from https://cas.cgiar.org/spia/publications/dna-fingerprinting-crop-varietal-identification-fit-purpose-protocols-their-costs

Rodriguez-Amaya, D. B., & Kimura, M. (2004). *Harvestplus handbook for carotenoid analysis* (Technical Report). Washington, DC and Cali: International Food Policy Research Institute (IFPRI) and International Center for Tropical Agriculture (CIAT).

Ruel, M. T., & Alderman, H. (2013). Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition? *The Lancet*, *382*(9891), 536–551. doi: 10.1016/S0140-6736(13)60843-0

Shankar, A. H., Genton, B., Semba, R. D., Baisor, M., Paino, J., Tamja, S., ... West, K. P. (1999). Effect of vitamin A supplementation on morbidity due to Plasmodium falciparum in young children in Papua New Guinea: A randomised trial. *Lancet (London, England)*, *354*(9174), 203–209. doi: 10.1016/S0140-6736(98)08293-2

Stephensen, C. B. (2001). Vitamin A, infection, and immune function. *Annual Review of Nutrition*, *21*, 167–192. doi: 10.1146/annurev.nutr.21.1.167

Stevenson, J., Gantier, M., Traxler, G., Kosmowski, F., & Macours, K. (2023). *The Challenge of Tracking the Reach of Post-Green Revolution Crop Breeding.* Research Square preprint. Retrieved from https://www.researchsquare.com/article/rs-3028333/v1 (In review) doi: 10.21203/rs.3.rs-3028333/v1

Thurnham, D. I., McCabe, G. P., Northrop-Clewes, C. A., & Nestel, P. (2003). Effects of subclinical infection on plasma retinol concentrations and assessment of prevalence of vitamin a deficiency: meta-analysis. *The Lancet*, *362*, 2052–2058. doi: 10.1016/S0140-6736(03)15099-4

Tumwegamire, S., Kapinga, R., Rubaihayo, P. R., LaBonte, D. R., Grüneberg, W. J., Burgos, G., & Mwanga, R. O. M. (2011). Starch, sucrose, β-carotene, iron, zinc, calcium, and magnesium in East African sweetpotato [Ipomoea batatas (L.) lam.] germplasm. *HortScience*, *46*(3), 348–357.

(UBOS), U. B. O. S. (2020). *Annual Agricultural Survey* (Tech. Rep.). Kampala, Uganda: Uganda Bureau of Statistics.

Uganda Bureau of Statistics, & ICF. (2018). *Uganda demographic and health survey 2016* (DHS Final Report No. FR333). Kampala, Uganda and Rockville, Maryland, USA: UBOS and ICF.

Uganda Bureau of Statistics, & ICF International Inc. (2012). *Uganda demographic and health survey 2011* (DHS Final Report No. FR264). Kampala, Uganda and Calverton, Maryland, USA: UBOS and ICF International Inc.

Uganda Bureau of Statistics, & ICF International Inc. (2013). *2011 Uganda DHS — addendum to chapter 11: Vitamin a* (Tech. Rep.). UBOS and ICF International Inc. (Addendum to DHS Final Report FR264; undated in the original)

UNICEF. (2020). *UNICEF Nutrition Strategy 2020-2030*. Author. Retrieved from https://www.unicef.org/media/92031/file/UNICEF%20Nutrition%20Strategy%202020-2030.pdf

World Health Organization. (2009). *Global prevalence of vitamin A deficiency in populations at risk 1995-2005 : WHO global database on vitamin A deficiency* (Tech. Rep.). Geneva: World Health Organization. Retrieved 2025-05-16, from https://iris.who.int/handle/10665/44110

World Health Organization. (2011). *Guideline : Vitamin A supplementation in infants and children 6-59 months of age* (Tech. Rep.). Geneva: World Health Organization. Retrieved 2025-05-16, from https://iris.who.int/handle/10665/44664

Wossen, T., Abay, K. A., & Abdoulaye, T. (2022). Misperceiving and misreporting input quality: Implications for input use and productivity. *Journal of Development Economics*, *157*, 102869. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304387822000414 doi: https://doi.org/10.1016/j.jdeveco.2022.102869

Wossen, T., Abdoulaye, T., Alene, A., Feleke, S., Menkir, A., & Manyong, V. (2019). Agricultural technology adoption and household welfare: Measurement and evidence. *Food Policy*, *87*, 101742. doi: 10.1016/j.foodpol.2019.101742

Wossen, T., Nguimkeu, P., Abdoulaye, T., Adebayo, A., Manyong, V., Feleke, S., & Tufa, A. (2019). Estimating the productivity impacts of technology adoption in the presence of misclassification. *American Journal of Agricultural Economics*, *101*(1), 1–16. doi: 10.1093/ajae/aay028

Yirga, C., Negisho, K., Oruko, L., Alemu, D., & Taxler, G. (2016). *Tracking the diffusion of crop varieties using DNA fingerprinting* (Tech. Rep.). Ethiopian Institute of Agricultural Research (EIAR).

10 Figures

Figure 1: Share of households per village with at least one orange/yellow fleshed or virus resistant variety.

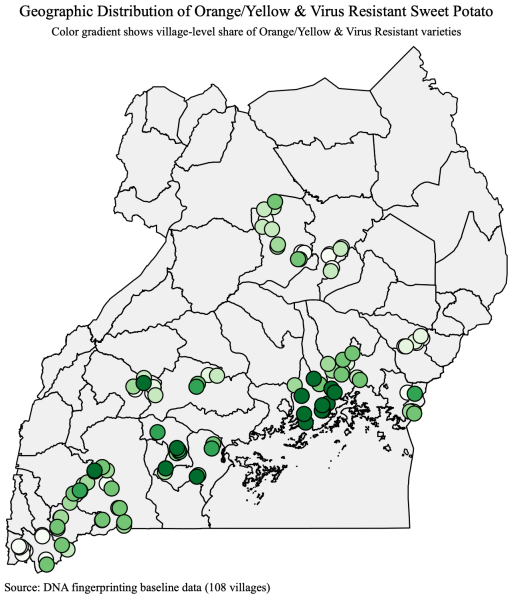


Figure 2: Share of households per village with at least one virus-resistant variety

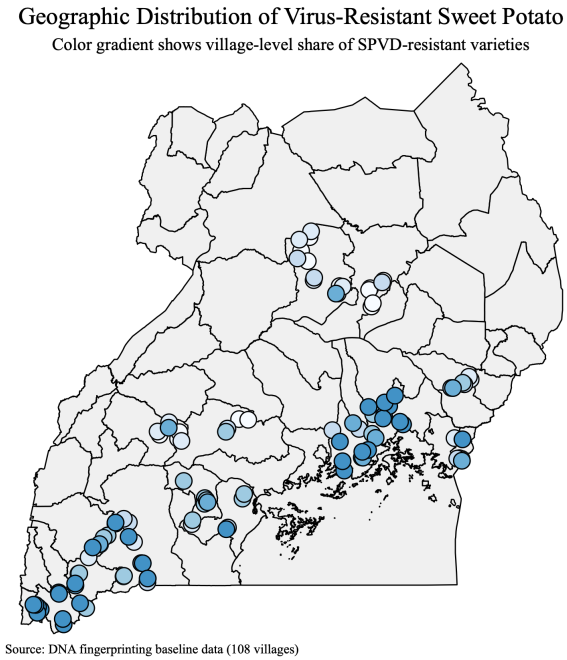
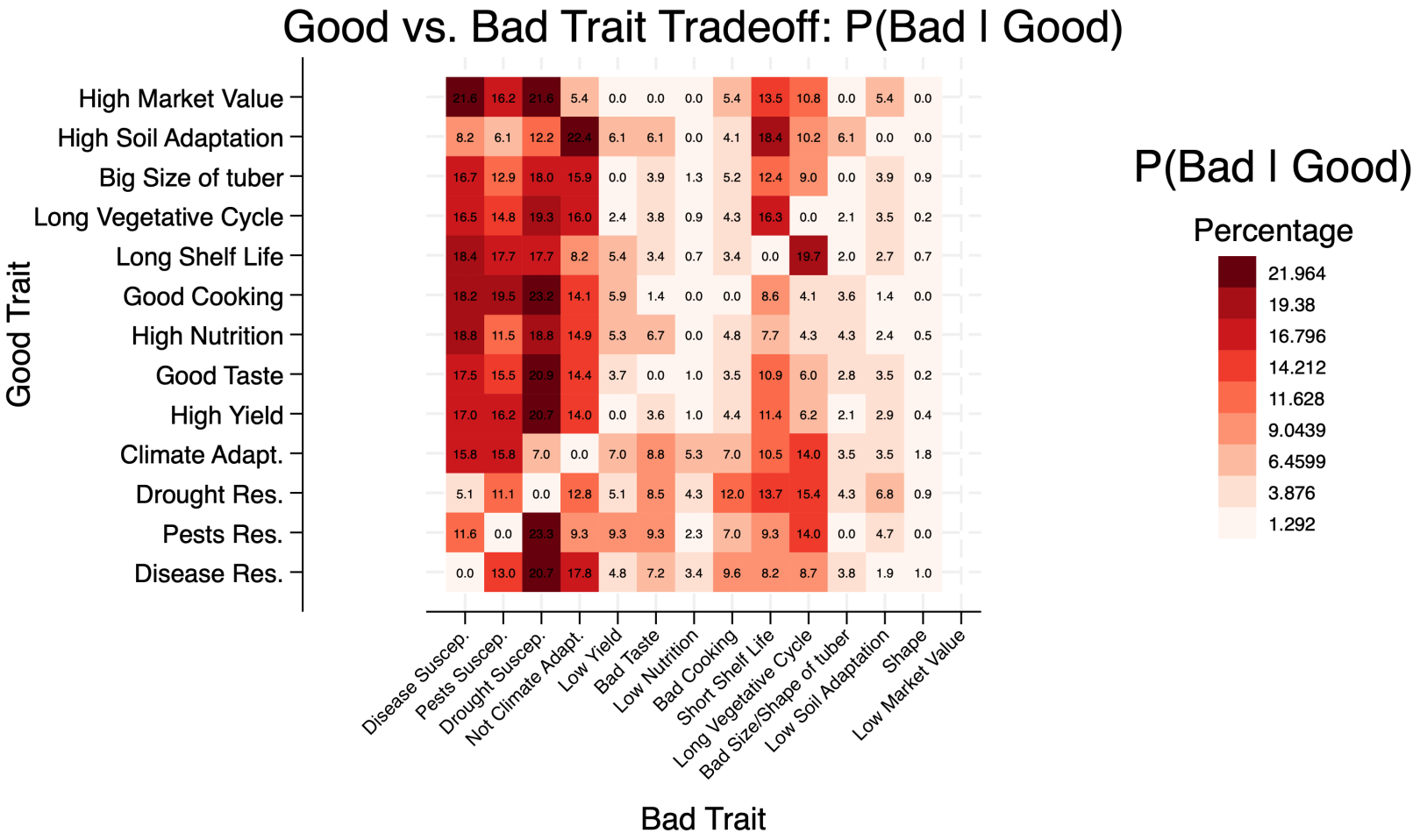
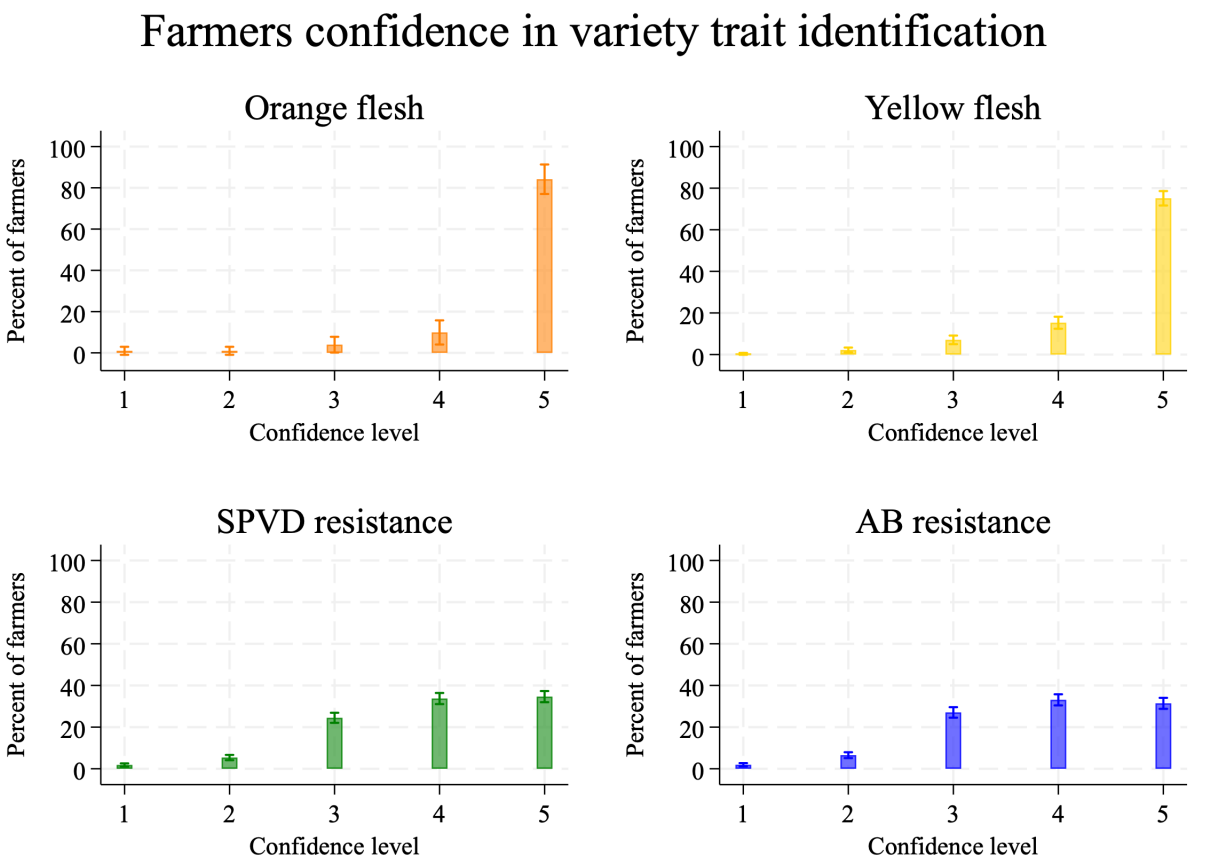


Figure 3: Conditional probability of "bad" trait based on "good" trait



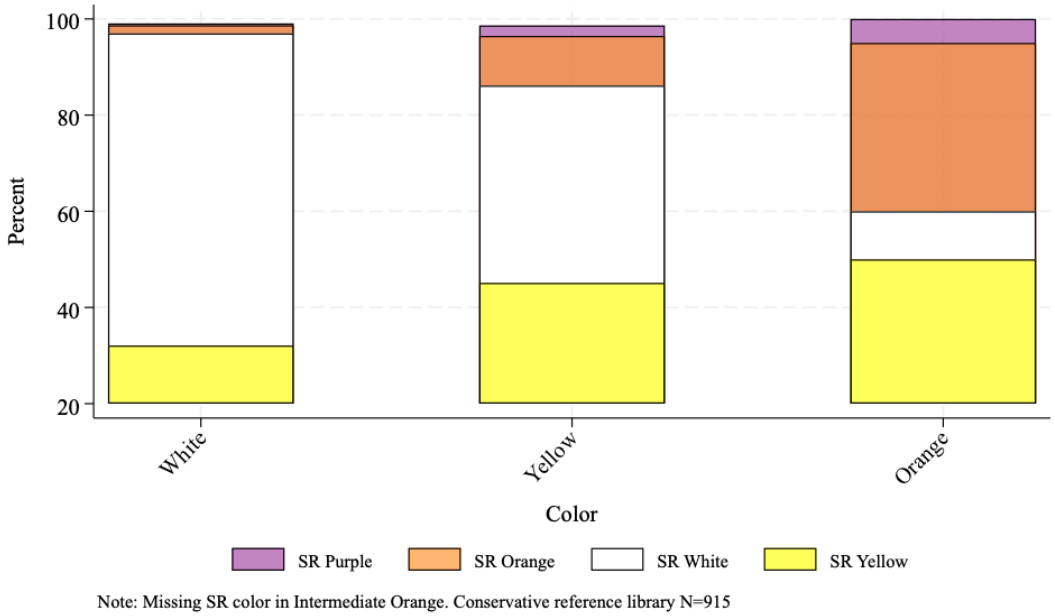
Notes: This figure displays conditional probabilities of reporting a "bad" trait (rows) given that a "good" trait is reported (columns). Each cell reports the conditional probability P(bad trait | good trait) computed from farmer responses in the control group at endline. Darker shading indicates a higher conditional probability.

Figure 4: Self-Reported Confidence in Assessments of Varietal Traits (Control Group, Endline).



Notes: Using endline data from the control group, the figure reports the distribution of farmers' stated confidence in their assessment of whether one of their own varieties possesses a given trait. Confidence is measured on a five-point Likert scale from 1 (not confident at all) to 5 (absolutely certain). For each trait, bars show the share of farmers selecting each confidence level. Error bars denote 95% confidence intervals.

Figure 5: Self-Reported Versus Objectively Measured Sweet Potato Flesh Color.



Notes: The figure compares farmers' self-reported sweet potato flesh color categories with an objective classification of flesh color at the variety level. For each objectively measured variety (horizontal axis), bars report the distribution of self-reported color categories among farmers who reported growing that variety.

11 Tables

Table 1: Sweet Potato Variety Adoption at Baseline

Panel A: Share of households with at least one variety		
Variety Type	Share	N
Orange-fleshed	0.062	1000
Yellow-fleshed	0.301	1000
White-fleshed	0.350	1000
Purple-fleshed	0.000	1000
SPVD-resistant	0.617	1000
AB-resistant	0.543	1000
Virus-resistant	0.620	1000
Yellow & Virus-resistant	0.222	1000
Identified	0.783	1000
Not Identified	0.820	1000
Total	1.000	1000
Panel B: Number of varieties per household		
Variety Type	Mean (SD)	N
Orange-fleshed	0.071 (0.290)	1000
Yellow-fleshed	0.429 (0.756)	1000
White-fleshed	0.533 (0.911)	1000
Purple-fleshed	0.000 (0.000)	1000
SPVD-resistant	1.025 (1.116)	1000
AB-resistant	0.793 (0.921)	1000
Virus-resistant	1.038 (1.129)	1000
Yellow & Virus-resistant	0.289 (0.605)	1000
Identified	1.641 (1.419)	1000
Not Identified	1.675 (1.323)	1000
Total	3.313 (1.838)	1000

Panel C: Share with at least one variety by region				
Variety Type	Eastern	Western	Northern	Central
Orange-fleshed	0.087	0.026	0.033	0.110
Yellow-fleshed	0.102	0.231	0.160	0.629
White-fleshed	0.607	0.362	0.110	0.309
Purple-fleshed	0.000	0.000	0.000	0.000
SPVD-resistant	0.719	0.704	0.293	0.647
AB-resistant	0.571	0.678	0.238	0.551
Virus-resistant	0.730	0.707	0.293	0.647
Yellow & Virus-resistant	0.061	0.185	0.083	0.478
Identified	0.908	0.838	0.508	0.805

Notes: The unit of observation is the household. Panel A reports the share of households cultivating at least one sweet potato variety in each category. Panel B reports the mean number of cultivated varieties per household in each category, with standard deviations in parentheses. Panel C reports the share of households cultivating at least one variety in each category by region. Variety categories are defined by objectively measured flesh color and resistance traits. "Yellow & Virus-resistant" indicates households cultivating at least one yellow-fleshed variety that is also virus-resistant. "Identified" ("Not identified") indicates households cultivating at least one variety that could (could not) be objectively identified. $N = 1{,}000$ households.

Table 2: Comparison Between Self-Reported Traits and DNA-Fingerprinting Results.

Variety Type	True Positive	False Positive	False Negative	True Negative
Orange-fleshed	0.027	0.045	0.009	0.918
Yellow-fleshed	0.142	0.284	0.261	0.311
SPVD-resistant	0.622	0.108	0.241	0.028
AB-resistant	0.444	0.277	0.223	0.056

Notes: The unit of observation is a farmer–variety. Using endline data from the control group, the table reports the share of farmer–variety observations in each cell of a 2×2 validation matrix comparing farmers' self-reported variety traits to DNA-fingerprinting results ($N = 671$ farmer–variety observations). "True Positive" indicates that the farmer reports the trait and DNA confirms the trait is present. "False Positive" indicates that the farmer reports the trait but DNA indicates the trait is absent. "False Negative" indicates that the farmer does not report the trait but DNA confirms the trait is present. "True Negative" indicates that the farmer does not report the trait and DNA indicates the trait is absent.

Table 3: Intent-to-Treat Effects on Recall of Sweet Potato Information.

	(1) Color→ Nutrition	(2) Orange = Best	(3) SPVD: Symptoms	(4) SPVD: Control	(5) AB: Symptoms	(6) AB: Control
T1	0.164***	0.123***	1.068***	0.765***	0.801***	0.615***
	(0.031)	(0.027)	(0.159)	(0.129)	(0.179)	(0.109)
T2	0.162***	0.148***	0.719***	0.734***	0.833***	0.531***
	(0.029)	(0.025)	(0.134)	(0.127)	(0.159)	(0.114)
Obs.	2105	2105	2105	2105	2105	2105
Control Mean	0.59	0.19	-0.00	-0.00	-0.00	-0.00
p-value: T1=T2	0.92	0.27	0.04	0.82	0.85	0.42

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on respondents' recall of key messages. The unit of observation is the respondent. "Color→ Nutrition" is an indicator equal to 1 if the respondent reports that sweet potato flesh color signals nutritional content. "Orange = Best" is an indicator equal to 1 if the respondent identifies orange-fleshed varieties as the most nutritious. Columns (3)–(6) report standardized indices of knowledge about symptoms and control practices for sweet potato virus disease (SPVD) and Alternaria blight (AB), constructed to have mean 0 and standard deviation 1 in the control group. All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the double-selection LASSO procedure. Standard errors are clustered at the village level. The "Control Mean" row reports the control-group mean of the dependent variable. The row "p-value: T1=T2" reports the p-value from a test of equality of the treatment coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Intent-to-Treat Effects on Correct Identification of Varietal Traits (DNA-Verified True Positives).

	(1) Orange	(2) Yellow	(3) SPVD res.	(4) AB res.
T1	-0.0182**	0.00634	-0.0480	-0.0349
	(0.01)	(0.03)	(0.04)	(0.05)
T2	-0.00142	0.0946***	-0.00423	0.0890*
	(0.01)	(0.03)	(0.04)	(0.05)
Obs.	2117	1503	1578	1578
Mean Control: TP	0.03	0.14	0.62	0.44
Mean Control: FP	0.04	0.28	0.11	0.28
Mean Control: FN	0.01	0.26	0.24	0.22
Mean Control: TN	0.92	0.31	0.03	0.06
P-value T1=T2	0.01	0.01	0.34	0.01

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on correct identification of varietal traits. The unit of observation is a farmer–variety. In each column, the dependent variable is an indicator equal to 1 if the farmer reports that the variety has the trait listed in the column heading and DNA fingerprinting confirms that the trait is present (a true positive), and 0 otherwise. The rows "Mean Control: TP," "Mean Control: FP," "Mean Control: FN," and "Mean Control: TN" report the corresponding shares of true positives, false positives, false negatives, and true negatives in the control group. All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. The row "p-value T1=T2" reports the p-value from a test of equality of the treatment coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Intent-to-Treat Effects on the Precision of Trait–Variety Matching.

	Share of TP				
	(1)	(2)	(3)	(4)	(5)
	Orange	Yellow	SPVD res.	AB res.	Yellow & Virus res.
T1	-0.0239	0.00159	-0.0155	0.000270	0.00540
	(0.02)	(0.03)	(0.04)	(0.05)	(0.03)
T2	-0.00595	0.0806**	-0.0375	0.0259	0.0532*
	(0.02)	(0.03)	(0.04)	(0.04)	(0.03)
Obs.	1012	1012	1012	1012	1012
Mean Control	0.04	0.14	0.50	0.42	0.11
P-value T1=T2	0.13	0.01	0.59	0.55	0.15

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on the precision of matching traits to varieties. The unit of observation is the household. In each column, the dependent variable is the share of varieties with the trait listed in the column heading that the household correctly identifies as possessing that trait (true positives), based on DNA fingerprinting as the benchmark. "Yellow & Virus res." indicates varieties that are both yellow-fleshed and virus-resistant. All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. The "Mean Control" row reports the control-group mean of the dependent variable. The row "p-value T1=T2" reports the p-value from a test of equality of the treatment coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Intent-to-Treat Effects on Household Area Cultivated by Variety Traits (Square Meters).

	(1) Orange	(2) Yellow	(3) SPVD	(4) AB	(5) Yellow & Virus res.	(6) White	(7) Purple	(8) Not identified	(9) GPS total area (sqm)
T1	-2.726	-2.251	27.40	11.17	4.032	32.99	0.178	6.577	0.493
	(3.55)	(16.12)	(26.59)	(22.38)	(15.28)	(21.82)	(0.35)	(22.28)	(44.18)
T2	-0.0879	31.95*	49.86*	65.83**	37.42**	13.47	0.183	14.86	101.1*
	(3.35)	(17.01)	(29.82)	(27.16)	(15.48)	(23.78)	(0.32)	(25.19)	(51.84)
Obs.	1012	1012	1012	1012	1012	1012	1012	1012	1012
Control Mean	8.94	89.66	198.41	145.95	63.41	102.89	0.29	178.59	552.38
p-value: T1=0	0.442	0.889	0.303	0.618	0.792	0.131	0.611	0.768	0.991
RI p-value: T1=0	0.458	0.901	0.387	0.665	0.811	0.205	0.631	0.808	0.997
p-value: T2=0	0.979	0.060	0.094	0.015	0.016	0.571	0.564	0.555	0.051
RI p-value: T2=0	0.980	0.100	0.093	0.013	0.041	0.598	0.612	0.585	0.060

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on household cultivated area (in square meters) allocated to sweet potato varieties grouped by objectively measured traits. The unit of observation is the household. In columns (1)–(8), the dependent variable is the total household area cultivated with varieties in the category shown in the column heading; in column (9), the dependent variable is total sweet potato area measured by GPS ("GPS total area"). Plot area is measured using GPS devices, and area allocated to each variety category is computed as total plot area multiplied by the self-reported share of the plot planted with each variety. Land area measures are winsorized at the 99th percentile. All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. The rows "p-value: T1=0" and "p-value: T2=0" report two-sided p-values from conventional inference; "RI p-value" rows report randomization-inference p-values based on 1,000 iterations using the randomization-c procedure. Table A5 reports the same specifications using land area measures without winsorization. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Intent-to-Treat Effects on Sweet Potato Production by DNA-Verified Varietal Traits (Kilograms).

	Consumed (KG)			Harvested (KG)			Still in field (KG)			Total (KG)		
	(1) Orange	(2) Yellow	(3) Virus res.	(4) Orange	(5) Yellow	(6) Virus res.	(7) Orange	(8) Yellow	(9) Virus res.	(10) Orange	(11) Yellow	(12) Virus res.
T1	0.0988	7.192*	13.87	2.528	6.789	17.66	1.551	6.659	43.70	4.079	13.45	55.79
	(0.76)	(3.85)	(9.88)	(2.29)	(4.14)	(14.81)	(1.62)	(7.33)	(41.37)	(2.98)	(9.49)	(48.35)
T2	0.755	6.235**	21.73**	1.260	8.532**	25.05*	2.000	10.05	47.81	3.260	18.58*	61.44
	(0.79)	(3.09)	(9.65)	(1.42)	(3.86)	(13.42)	(2.01)	(7.20)	(32.18)	(2.89)	(9.77)	(37.86)
Obs.	1012	1012	1012	1012	1012	1012	1012	1012	1012	1012	1012	1012
Control Mean	0.79	9.33	43.59	0.79	9.77	50.36	1.79	26.73	135.78	2.59	36.50	186.14
P-value T1=T2	0.54	0.06	0.07	0.53	0.04	0.16	0.55	0.37	0.31	0.37	0.13	0.25

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on self-reported sweet potato quantities (in kilograms) by DNA-verified varietal traits. The unit of observation is the household. Panel columns report quantities consumed, harvested, still in the field at the time of the survey, and total quantity (sum of the three components), each disaggregated by varieties classified as orange-fleshed, yellow-fleshed, or virus-resistant based on DNA fingerprinting. All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. The "Control Mean" row reports the control-group mean of the dependent variable. The row "p-value T1=T2" reports the p-value from a test of equality of the treatment coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Intent-to-Treat Effects on Consumption of Orange- or Yellow-Fleshed Sweet Potato among Children under Age 2.

	Last 7 days		Last 6 months
	(1) Any Consumption	(2) Number of times consumed	(3) Consumed Once a week
T1	0.0418	-0.552	-0.0387
	(0.08)	(0.42)	(0.08)
T2	0.146**	0.414	0.106
	(0.07)	(0.49)	(0.08)
Obs.	202	202	202
Control Mean	0.46	2.37	0.56
p-value: T1=0	0.578	0.189	0.619
RI p-value: T1=0	0.630	0.259	0.659
p-value: T2=0	0.032	0.398	0.191
RI p-value: T2=0	0.093	0.408	0.230
p-value: T1=T2	0.229	0.055	0.109
RI p-value: Joint	0.214	0.158	0.227

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on consumption of orange- or yellow-fleshed sweet potato among children under age 2. The unit of observation is the child. Columns (1) and (2) use outcomes measured over the last 7 days: "Any Consumption" is an indicator equal to 1 if the child consumed orange- or yellow-fleshed sweet potato at least once, and "Number of times consumed" is the reported number of consumption occasions. Column (3) uses outcomes measured over the last 6 months: "Consumed Once a week" is an indicator equal to 1 if the child consumed orange- or yellow-fleshed sweet potato at least once per week on average. All specifications are estimated by OLS and include stratum fixed effects, child age fixed effects, and a child gender indicator; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. Rows labeled "p-value" report two-sided p-values from conventional inference; rows labeled "RI p-value" report randomization-inference p-values based on 1,000 iterations using the randomization-c procedure. The row "RI p-value: Joint" reports the randomization-inference p-value for a joint test that both treatment coefficients are equal to zero. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Intent-to-Treat Effects on Diarrhea Incidence and Duration among Children under Age 2.

	(1) Last 2 weeks	(2) Days last episode	(3) Last 6 months
T1	-0.0811	-0.865*	-0.0979
	(0.05)	(0.46)	(0.09)
T2	-0.165**	-1.266***	-0.159
	(0.07)	(0.40)	(0.10)
Obs.	216	216	216
Control Mean	0.32	2.92	0.65
p-value: T1=0	0.137	0.058	0.251
RI p-value: T1=0	0.341	0.110	0.349
p-value: T2=0	0.027	0.002	0.106
RI p-value: T2=0	0.048	0.013	0.113
RI p-value: Joint	0.114	0.051	0.285

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on diarrhea outcomes among children under age 2. The unit of observation is the child. Column (1) uses an indicator equal to 1 if the child had diarrhea in the last 2 weeks. Column (2) reports the number of days in the child's most recent diarrhea episode. Column (3) uses an indicator equal to 1 if the child had diarrhea at any point in the last 6 months. All specifications are estimated by OLS and include stratum fixed effects, child age fixed effects, and a child gender indicator; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. Rows labeled "p-value" report two-sided p-values from conventional inference; rows labeled "RI p-value" report randomization-inference p-values based on 1,000 iterations using the randomization-c procedure. The row "RI p-value: Joint" reports the randomization-inference p-value for a joint test that both treatment coefficients are equal to zero. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Intent-to-Treat Effects on Consumption of Orange- or Yellow-Fleshed Sweet Potato among Children under Age 2 in Community-Information Households.

	Last 7 days		Last 6 months
	(1)	(2)	(3)
	Incidence	Nb. of times	At Least Once/Week
T2 (G2)	0.171*	0.141	0.154
	(0.098)	(0.536)	(0.095)
Obs.	213	213	213
Control Mean	0.46	2.16	0.54

Notes: This table reports intent-to-treat estimates of the effects of receiving a community card in the T2 arm on consumption of orange- or yellow-fleshed sweet potato among children under age 2. The unit of observation is the child. Community-information households are those (and their children) that received a community card in T2 containing DNA-fingerprinting results about the varietal traits of individual-information households' sweet potato varieties (rather than DNA results for their own varieties). The reported coefficient ("T2 (G2)") compares outcomes for children in community-information households to outcomes in the comparison group consisting of the pure control group and the T1 group (which did not receive any card). Columns (1) and (2) use outcomes measured over the last 7 days: "Incidence" is an indicator equal to 1 if the child consumed orange- or yellow-fleshed sweet potato at least once, and "Nb. of times" is the reported number of consumption occasions. Column (3) uses outcomes measured over the last 6 months: "At Least Once/Week" is an indicator equal to 1 if the child consumed orange- or yellow-fleshed sweet potato at least once per week on average. All specifications are estimated by OLS and include stratum fixed effects, child age fixed effects, and a child gender indicator; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. The "Control Mean" row reports the mean of the dependent variable in the comparison group. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Intent-to-Treat Effects on Diarrhea Incidence and Duration among Children under Age 2 in Community-Information Households.

	(1) Last 2 weeks	(2) Days last episode	(3) Last 6 months
T2 (G2)	-0.189**	-0.147	-0.00985
	(0.09)	(0.53)	(0.10)
Obs.	226	226	226
Control Mean	0.31	2.45	0.60

Notes: This table reports intent-to-treat estimates of the effects of receiving a community card in the T2 arm on diarrhea outcomes among children under age 2. The unit of observation is the child. Community-information households are those (and their children) that received a community card in T2 containing DNA-fingerprinting results about the varietal traits of individual-information households' sweet potato varieties (rather than DNA results for their own varieties). The reported coefficient ("T2 (G2)") compares outcomes for children in community-information households to outcomes in the comparison group consisting of the pure control group and the T1 group (which did not receive any card). Column (1) uses an indicator equal to 1 if the child had diarrhea in the last 2 weeks. Column (2) reports the number of days in the child's most recent diarrhea episode. Column (3) uses an indicator equal to 1 if the child had diarrhea at any point in the last 6 months. All specifications are estimated by OLS and include stratum fixed effects, child age fixed effects, and a child gender indicator; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. The "Control Mean" row reports the mean of the dependent variable in the comparison group. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Intent-to-Treat Effects on Household Land Allocation by Self-Reported Varietal Traits (Square Meters).

	(1)	(2)	(3)	(4)	(5)
	Orange	Yellow	SPVD-resistant	AB-resistant	Orange/Yellow & SPVD/AB
T1	-2.856	23.64	0.820	-6.176	20.16
	(6.45)	(27.27)	(37.81)	(46.10)	(28.79)
T2	3.221	62.61**	69.24	69.06	56.48**
	(6.02)	(28.56)	(44.11)	(48.56)	(28.44)
Observations	1012	1012	1012	1012	1012
Mean Control	32.51	170.51	381.48	372.84	149.50

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on household land allocation (in square meters) to sweet potato varieties grouped by self-reported traits. The unit of observation is the household. In each column, the dependent variable is the total household area cultivated with varieties reported to have the trait(s) listed in the column heading; "Orange/Yellow & SPVD/AB" indicates varieties reported to be orange- or yellow-fleshed and resistant to either sweet potato virus disease (SPVD) or Alternaria blight (AB). All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 13: Intent-to-Treat Effects on Household Land Allocation by DNA-Verified Varietal Traits and Baseline Varieties without Trade-Offs (Square Meters).

	(1) Orange	(2) Yellow	(3) SPVD	(4) AB	(5) Yellow & Virus res.	(6) White	(7) Purple	(8) Not identified	(9) GPS total area (sqm)
T1	-2.508	-2.921	29.63	9.820	7.309	28.98	0.407	17.46	3.155
	(3.50)	(12.48)	(25.13)	(19.46)	(10.72)	(23.46)	(0.44)	(25.16)	(39.58)
T1*Orange/Yellow & SPVD/AB	-4.483	-12.16	-53.65	-10.70	-20.15	-23.08	-1.067*	-78.32*	-43.67
	(8.33)	(33.07)	(44.84)	(37.92)	(33.06)	(33.89)	(0.55)	(44.96)	(81.66)
T2	1.431	5.554	5.098	22.95	13.42	-3.661	-0.166	32.25	51.70
	(3.31)	(14.33)	(25.17)	(22.70)	(12.02)	(20.31)	(0.37)	(28.06)	(45.83)
T2*Orange/Yellow & SPVD/AB	-4.780	85.26**	167.2***	141.2**	65.82*	95.61**	1.188	-58.71	155.9
	(8.76)	(35.18)	(64.74)	(58.48)	(33.94)	(42.16)	(1.10)	(46.98)	(102.31)
Orange/Yellow & SPVD/AB	11.97*	-103.6**	-115.6*	-96.08*	-84.30**	-29.85	0.366	35.93	-259.2*
	(7.20)	(45.29)	(67.88)	(54.71)	(37.97)	(42.96)	(0.85)	(48.46)	(136.92)
Observations	1012	1012	1012	1012	1012	1012	1012	1012	1012
Mean Control	6.05	60.42	175.85	130.17	37.49	103.41	0.33	170.08	540.06
p-value: T1=0	0.473	0.815	0.238	0.614	0.495	0.217	0.355	0.488	0.936
p-value: T2=0	0.666	0.698	0.839	0.312	0.264	0.857	0.656	0.250	0.259
P-value T1+T1xX=0	0.372	0.677	0.589	0.981	0.710	0.842	0.060	0.130	0.640
P-value T2+T2xX=0	0.684	0.009	0.006	0.004	0.020	0.029	0.278	0.549	0.046

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on household land allocation (in square meters) to sweet potato varieties grouped by DNA-verified traits, allowing effects to vary with the presence of baseline varieties without trade-offs at the household level. The unit of observation is the household. In columns (1)–(8), the dependent variable is the total household area cultivated with varieties in the category shown in the column heading; in column (9), the dependent variable is total sweet potato area measured by GPS ("GPS total area"). Plot area is measured using GPS devices, and area allocated to each trait category is computed as total plot area multiplied by the self-reported share of the plot planted with each variety. "Orange/Yellow & SPVD/AB" is an indicator equal to 1 if the household cultivates at least one baseline variety that is orange- or yellow-fleshed and resistant to either sweet potato virus disease (SPVD) or Alternaria blight (AB), based on DNA fingerprinting. Rows labeled "T1*Orange/Yellow & SPVD/AB" and "T2*Orange/Yellow & SPVD/AB" report interactions between treatment indicators and this baseline indicator. The "Mean Control" row reports the mean of the dependent variable in the control group among households with Orange/Yellow & SPVD/AB = 0. Rows "p-value: T1=0" and "p-value: T2=0" report two-sided p-values testing the main treatment effects; rows "p-value T1+T1xX=0" and "p-value T2+T2xX=0" report two-sided p-values testing the sum of the main effect and interaction term (where X denotes Orange/Yellow & SPVD/AB). All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 14: Intent-to-Treat Effects on Household Land Allocation by DNA-Verified Varietal Traits and Village-Level Baseline Varieties without Trade-Offs (Square Meters).

	(1) Orange	(2) Yellow	(3) SPVD	(4) AB	(5) Yellow & Virus res.	(6) White	(7) Purple	(8) Not identified	(9) GPS total area (sqm)
T1	-0.657	9.641	46.63	13.29	23.19	54.14	0.0530	77.76**	0.111
	(3.18)	(17.78)	(48.55)	(43.52)	(19.46)	(36.37)	(0.31)	(34.07)	(68.68)
T1*Orange/Yellow & SPVD/AB - village	-2.984	-13.94	-23.97	-1.009	-22.47	-27.67	0.102	-85.46**	8.263
	(5.34)	(24.34)	(54.87)	(46.24)	(25.72)	(44.33)	(0.47)	(40.73)	(84.86)
T2	0.611	-8.760	15.50	4.402	0.991	41.99*	0.385	105.7***	55.35
	(3.35)	(12.95)	(44.06)	(43.17)	(13.85)	(24.54)	(0.49)	(35.10)	(67.46)
T2*Orange/Yellow & SPVD/AB - village	-0.258	43.85*	38.36	71.77	38.76	-42.08	-0.302	-113.3**	35.77
	(5.81)	(24.89)	(58.11)	(52.35)	(23.62)	(43.19)	(0.61)	(50.34)	(94.07)
Orange/Yellow & SPVD/AB - village	4.220	17.46	9.648	-23.96	9.718	53.17*	0.684*	11.81	-58.38
	(3.94)	(17.60)	(47.09)	(43.25)	(18.17)	(30.49)	(0.40)	(29.85)	(65.66)
Observations	1012	1012	1012	1012	1012	1012	1012	1012	1012
Mean Control	11.71	116.41	222.79	151.24	81.36	130.05	0.40	195.32	593.12
P-value T1+T1xX=0	0.41	0.81	0.44	0.59	0.97	0.30	0.71	0.76	0.87
P-value T2+T2xX=0	0.94	0.09	0.14	0.01	0.03	1.00	0.82	0.82	0.15

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on household land allocation (in square meters) to sweet potato varieties grouped by DNA-verified traits, allowing effects to vary with the presence of baseline varieties without trade-offs at the village level. The unit of observation is the household. In columns (1)–(8), the dependent variable is the total household area cultivated with varieties in the category shown in the column heading; in column (9), the dependent variable is total sweet potato area measured by GPS ("GPS total area"). Plot area is measured using GPS devices, and area allocated to each trait category is computed as total plot area multiplied by the self-reported share of the plot planted with each variety. "Orange/Yellow & SPVD/AB – village" is an indicator equal to 1 if the village has at least one baseline variety that is orange- or yellow-fleshed and resistant to either sweet potato virus disease (SPVD) or Alternaria blight (AB), based on DNA fingerprinting. Rows labeled "T1*Orange/Yellow & SPVD/AB – village" and "T2*Orange/Yellow & SPVD/AB – village" report interactions between treatment indicators and this village-level baseline indicator. The "Mean Control" row reports the mean of the dependent variable in the control group among households in villages with Orange/Yellow & SPVD/AB – village = 0. Rows "p-value T1+T1xX=0" and "p-value T2+T2xX=0" report two-sided p-values testing the sum of the main effect and interaction term (where X denotes Orange/Yellow & SPVD/AB – village). All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 15: Intent-to-Treat Effects on Household Land Allocation by Baseline Village Varietal Shares (Terciles).

	(1) Orange	(2) Yellow	(3) SPVD	(4) AB	(5) Yellow & Virus res.	(6) White	(7) Purple	(8) Not identified	(9) GPS total area (sqm)
T1 - 1st tercile	-8.120**	-9.774	-0.888	-52.03**	-9.096	7.573	0.558	-38.82	-171.4***
	(3.653)	(18.287)	(30.473)	(20.475)	(19.488)	(24.864)	(0.595)	(32.318)	(48.107)
T1 - 2nd tercile	-4.053	-1.130	29.87	16.82	6.024	-8.800	0.684	-2.903	-22.29
	(5.055)	(14.112)	(35.961)	(24.496)	(14.621)	(22.809)	(0.626)	(39.395)	(46.477)
T1 - 3rd tercile	-1.367	-0.435	67.19	32.22	6.915	86.95**	-0.289	47.08	135.5*
	(6.567)	(17.202)	(42.551)	(26.005)	(20.403)	(37.689)	(0.177)	(37.466)	(79.003)
T2 - 1st tercile	-7.060*	-30.67	-51.61	-47.83*	-24.85	-1.262	0.236	-10.68	-79.12
	(3.752)	(24.965)	(35.190)	(24.830)	(17.019)	(23.362)	(0.528)	(44.376)	(77.469)
T2 - 2nd tercile	-5.091	-15.29	-3.415	-12.34	-19.72	15.79	-0.289	6.836	36.31
	(4.201)	(21.577)	(35.963)	(29.344)	(12.096)	(29.540)	(0.177)	(55.513)	(83.909)
T2 - 3rd tercile	3.076	118.5***	168.1***	193.6***	116.1***	32.13	0.425	-5.852	269.7***
	(6.729)	(41.191)	(59.892)	(59.801)	(38.648)	(32.915)	(0.694)	(37.896)	(90.189)
Observations	1012	1012	1012	1012	1012	1012	1012	1012	1012
Mean Control	8.94	89.66	198.41	145.95	63.41	102.89	0.29	178.59	552.38
P-value T1(1st)+T1(2nd)+T1(3rd)=0	0.262	0.782	0.221	0.956	0.922	0.158	0.332	0.945	0.645
P-value T1(1st)=T1(2nd)	0.333	0.606	0.413	0.004	0.461	0.554	0.879	0.417	0.005
P-value T1(1st)=T1(3rd)	0.255	0.627	0.144	0.001	0.531	0.056	0.136	0.041	0.000
P-value T1(2nd)=T1(3rd)	0.697	0.955	0.439	0.599	0.965	0.013	0.105	0.301	0.049
P-value T2(1st)+T2(2nd)+T2(3rd)=0	0.433	0.230	0.217	0.093	0.145	0.464	0.707	0.916	0.164
P-value T2(1st)=T2(2nd)	0.543	0.593	0.232	0.263	0.748	0.581	0.291	0.792	0.283
P-value T2(1st)=T2(3rd)	0.116	0.001	0.001	0.000	0.001	0.310	0.821	0.927	0.002
P-value T2(2nd)=T2(3rd)	0.224	0.002	0.009	0.001	0.000	0.663	0.287	0.838	0.038

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on household land allocation (in square meters) to sweet potato varieties grouped by DNA-verified traits, allowing treatment effects to vary across terciles of baseline village varietal shares. The unit of observation is the household. Each column reports the total household area cultivated with varieties in the category shown in the column heading; the final column reports total sweet potato area measured by GPS ("GPS total area"). Rows labeled "T1 – 1st tercile" through "T1 – 3rd tercile" and "T2 – 1st tercile" through "T2 – 3rd tercile" report treatment effects estimated for households in villages in the corresponding tercile of the baseline distribution of the relevant village-level varietal share. All specifications are estimated by OLS and include region fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. Rows labeled "p-value" report two-sided p-values from tests of joint significance across terciles and equality of coefficients across terciles, as indicated. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 16: Intent-to-Treat Effects on Household Area Cultivated by Variety Traits: Heterogeneity by Female Respondent (Square Meters).

	(1) Orange	(2) Yellow	(3) SPVD	(4) AB	(5) Yellow & Virus res.	(6) White	(7) Purple	(8) Not identified	(9) GPS total area (sqm)
T1	-4.560	-20.85	20.33	-2.181	-0.219	26.64	0.992	-31.95	-57.26
	(6.32)	(21.73)	(41.26)	(33.64)	(21.71)	(31.25)	(0.74)	(31.76)	(52.77)
T1*Fem.Resp.	2.595	23.27	11.96	21.57	1.764	12.03	-1.151	52.24	86.83
	(6.56)	(23.55)	(42.74)	(32.94)	(20.41)	(33.41)	(0.86)	(33.18)	(67.51)
T2	-3.366	1.227	16.19	49.28	16.22	-8.550	0.265	-11.88	109.0
	(6.38)	(24.40)	(44.20)	(43.08)	(19.97)	(29.62)	(0.41)	(37.43)	(102.06)
T2*Fem.Resp.	4.863	46.15	49.21	24.07	32.05	32.00	-0.132	40.11	-10.27
	(7.00)	(29.13)	(47.54)	(42.91)	(26.19)	(29.15)	(0.68)	(33.91)	(107.75)
Observations	1012	1012	1012	1012	1012	1012	1012	1012	1012
Mean Control	12.26	123.66	217.39	163.44	82.29	120.10	0.00	199.47	605.47
P-value T1+T1xFem.Resp.=0	0.60	0.89	0.26	0.41	0.92	0.12	0.70	0.42	0.59
P-value T2+T2xFem.Resp.=0	0.68	0.02	0.05	0.01	0.01	0.37	0.78	0.27	0.06

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on household cultivated area (in square meters) allocated to sweet potato varieties grouped by objectively measured traits, allowing treatment effects to vary by whether the survey respondent is female. The unit of observation is the household. In columns (1)–(8), the dependent variable is the total household area cultivated with varieties in the category shown in the column heading; in column (9), the dependent variable is total sweet potato area measured by GPS ("GPS total area"). Plot area is measured using GPS devices, and area allocated to each variety category is computed as total plot area multiplied by the self-reported share of the plot planted with each variety. "Fem. Resp." is an indicator equal to 1 if the survey respondent is female, and the interaction terms "T1×Fem. Resp." and "T2×Fem. Resp." allow treatment effects to vary by respondent gender. Land area measures are winsorized at the 99th percentile. All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. Rows "p-value T1+T1×Fem. Resp.=0" and "p-value T2+T2×Fem. Resp.=0" report two-sided p-values testing the sum of the main treatment effect and the interaction term. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 17: Intent-to-Treat Effects on Respiratory and Febrile Symptoms among Children under Age 2.

	Last 2 weeks				Last 6 months			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Diff. breathing	Cough	Fever	Malaria	Diff. breathing	Cough	Fever	Malaria
T1	-0.0308	0.0908	-0.0918	-0.0789	-0.0832	-0.0368	-0.0109	0.0675
	(0.056)	(0.060)	(0.066)	(0.068)	(0.067)	(0.052)	(0.048)	(0.066)
T2	-0.0151	-0.102	-0.162**	0.0327	-0.0240	-0.0982	-0.0266	0.0428
	(0.064)	(0.080)	(0.077)	(0.089)	(0.081)	(0.064)	(0.058)	(0.083)
Obs.	216	216	216	216	216	216	216	216
Control Mean	0.29	0.71	0.65	0.44	0.34	0.92	0.75	0.62
p-value: T1=T2	0.82	0.02	0.38	0.24	0.50	0.40	0.80	0.78

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on respiratory and febrile symptoms among children under age 2. The unit of observation is the child. Each dependent variable is an indicator equal to 1 if the caregiver reports that the child experienced the symptom listed in the column heading over the reference period shown in the panel header (last 2 weeks or last 6 months). "Diff. breathing" indicates difficulty breathing. All specifications are estimated by OLS and include stratum fixed effects, child age fixed effects, and a child gender indicator; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. The row "p-value: T1=T2" reports the p-value from a test of equality of the treatment coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 18: Intent-to-Treat Effects on Reported Night Vision Difficulties among Children under Age 16.

	(1) Difficulty seeing at night	(2) Difficulty seeing at night and wears glasses
T1	-0.00626	0.000448
	(0.05)	(0.04)
T2	-0.0361	0.00948
	(0.05)	(0.04)
Obs.	606	606
Control Mean	0.69	0.17
P-value T1=T2	0.61	0.82

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on reported night vision difficulties among children under age 16. The unit of observation is the child. In column (1), the dependent variable is an indicator equal to 1 if the caregiver reports that the child has difficulty seeing at night. In column (2), the dependent variable is an indicator equal to 1 if the caregiver reports that the child has difficulty seeing at night and wears glasses. All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. The row "p-value T1=T2" reports the p-value from a test of equality of the treatment coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 19: Intent-to-Treat Effects on Diarrhea Outcomes among Children under Age 2: Heterogeneity by Nutrition Training.

	(1)	(2)	(3)
	Last 2 weeks	Days last episode	Last 6 months
T1	-0.0320	-0.810	-0.0764
	(0.08)	(0.55)	(0.09)
T1xNutr. Training	-0.161	0.257	-0.0870
	(0.17)	(0.91)	(0.17)
T2	-0.135	-1.487***	-0.159
	(0.10)	(0.55)	(0.12)
T2xNutr. Training	-0.108	0.274	-0.0286
	(0.17)	(0.77)	(0.18)
Nutr. Training	0.0532	-1.508**	-0.0886
	(0.10)	(0.76)	(0.12)
Obs.	216	216	216
Control Mean	0.32	2.92	0.65
p-value: T1+T1xNutr.Training=0	0.13	0.34	0.29
p-value: T2+T2xNutr.Training=0	0.04	0.02	0.17

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on diarrhea outcomes among children under age 2, allowing effects to differ by whether the household received nutrition training. The unit of observation is the child. Column (1) uses an indicator equal to 1 if the child had diarrhea in the last 2 weeks. Column (2) reports the number of days in the child's most recent diarrhea episode. Column (3) uses an indicator equal to 1 if the child had diarrhea at any point in the last 6 months. "Nutr. Training" is an indicator equal to 1 if the child's household received nutrition training, and the interaction terms "T1×Nutr. Training" and "T2×Nutr. Training" allow treatment effects to vary by training status. All specifications are estimated by OLS and include stratum fixed effects, child age fixed effects, and a child gender indicator; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. The rows "p-value: T1+T1×Nutr. Training=0" and "p-value: T2+T2×Nutr. Training=0" report two-sided p-values testing the sum of the main treatment effect and the interaction term. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A Appendix: Model and Derivations

This appendix sets out the model of Section 3 in full, solves the farmer's problem, and derives the five predictions. Each proposition corresponds to the prediction with the same number.

A. Setup

Varieties, land, and traits. Farmer i holds a set of varieties $\mathcal{V}_i$ and allocates land $a_i(v) \geq 0$ to each. Total sweet potato area is $A_i = \sum_{v \in \mathcal{V}_i} a_i(v)$. The intervention targets two traits, $T = \{c, r\}$: $x_c(v)$, the vitamin A content of the root, which flesh color signals, and $x_r(v)$, resistance to disease.[24] I treat these as continuous indices, though in the data they are binary, and measure them relative to a reference variety, so that $x_k(v) = 0$ denotes the reference level. All other attributes of a variety are untargeted: the experiment provided no direct information about them.

Value. Farmers care about what a variety delivers (nutrition, the harvest, taste), and these preferences over outcomes are fixed. The value to farmer i of a unit of land planted with variety v is

$$\pi_i(v) = \bar{\pi} + \sum_{k \in T} \omega_{ik}\, x_k(v) + \eta_i(v), \tag{A.1}$$

where $\bar{\pi}$ is the value of a reference variety, $\omega_{ik} \geq 0$ is the value to her of the outcomes that trait k delivers, and $\eta_i(v)$ collects the value of everything else about the variety.[25]

Cues and beliefs. The farmer does not observe $x_k(v)$. For each targeted trait she observes a visual cue in her own field,

$$s_{ik}(v) = x_k(v) + \varepsilon_{ik}(v), \qquad \mathbb{E}[\varepsilon_{ik}(v)] = 0, \quad \mathrm{Var}[\varepsilon_{ik}(v)] = \sigma_k^2, \tag{A.2}$$

where the noise arises in how the cue is judged in the field. Her belief combines a prior with the cue, as in (1): $\tilde{x}_{ik}(v) = (1 - \rho_{ik})\, \tilde{x}^0_{ik}(v) + \rho_{ik}\, s_{ik}(v)$.

Perceived trade-offs. The parameter $\tau_i \in [0, 1]$ measures how strongly the farmer believes that desirable traits trade off against one another. It enters in two ways. First,

[24] The experiment covers two diseases, SPVD and AB. I write a single resistance trait to keep the notation light.

[25] The linear form should be read as a first-order approximation around the farmer's existing portfolio. Complementarities between traits would strengthen, rather than reverse, the predicted reallocations toward varieties that bundle desirable targeted traits.

she associates trait k with costs $\tau_i\,\kappa_{ik}$ for other outcomes, $\kappa_{ik} \geq 0$, so that perceived value is, as in (2),

$$\tilde{\pi}_i(v) = \bar{\pi} + \sum_{k \in T} \big(\omega_{ik} - \tau_i\,\kappa_{ik}\big)\,\tilde{x}_{ik}(v) + \eta_i(v).$$

Second, her prior about a variety's resistance may be lower, by $\tau_i\,\theta\,\tilde{x}_{ic}(v)$ with $\theta \geq 0$, the more she regards it as rich in vitamin A. Let $\hat{x}_{ir}(v)$ denote the belief about resistance she would hold without this inference, formed from her prior and the resistance cue as in (1). Since the prior carries weight $1 - \rho_{ir}$,

$$\tilde{x}_{ir}(v) = \hat{x}_{ir}(v) - (1 - \rho_{ir})\,\tau_i\,\theta\,\tilde{x}_{ic}(v). \tag{A.3}$$

I write the inference in this direction to keep the algebra simple; a remark after Proposition 3 lets it run in both directions.

B. The farmer's problem

The farmer chooses the area of each variety she holds to maximize

$$U_i = \sum_{v \in \mathcal{V}_i} \Big[a_i(v)\,\tilde{\pi}_i(v) - \frac{\phi_i}{2}\,a_i(v)^2\Big], \qquad a_i(v) \geq 0, \tag{A.4}$$

where $\phi_i > 0$ governs how fast the net return to a variety falls as it is given more land. The objective is strictly concave and separable across varieties, so the solution is unique and given variety by variety by the first-order condition $\tilde{\pi}_i(v) - \phi_i\,a_i(v) = 0$:

$$a_i(v) = \frac{\tilde{\pi}_i(v)}{\phi_i} \;\text{ if } \tilde{\pi}_i(v) > 0, \qquad a_i(v) = 0 \;\text{ otherwise.} \tag{A.5}$$

Three implications follow. For a variety she plants, a change in perceived value changes its area by

$$\Delta a_i(v) = \frac{\Delta\tilde{\pi}_i(v)}{\phi_i}. \tag{A.6}$$

Total area is $A_i = \sum_v \tilde{\pi}_i(v)/\phi_i$ over the varieties she plants. And the value she obtains from a variety at the optimum is $\tilde{\pi}_i(v)^2/(2\phi_i)$, so that if introducing a variety she does not hold costs $F_i(v) > 0$, she introduces it only if

$$\frac{\tilde{\pi}_i(v)^2}{2\phi_i} > F_i(v). \tag{A.7}$$

With correct beliefs and $\tau_i = 0$, $\tilde{\pi}_i(v) = \pi_i(v)$ and (A.5) is the allocation she would choose with full information.[26]

C. Treatments

Superscripts C, $T1$, and $T2$ denote the control group and the two treatment arms, and $\Delta^g y = y^g - y^C$ is the difference between arm g and control for the same farmer and variety, holding fixed her priors, the cues, and ω_{ik}, κ_{ik}, ϕ_i, and $\eta_i(v)$.

T1 raises the weight on the cues, $\rho_{ik}^{T1} > \rho_{ik}^{C}$ for $k \in T$, and leaves the cues and τ_i unchanged. I write $\Delta\rho_{ik} = \rho_{ik}^{T1} - \rho_{ik}^{C}$.

T2 delivers the same session, so $\rho_{ik}^{T2} = \rho_{ik}^{T1}$. For each DNA-matched variety of a sampled farmer the card states both traits, so that $\tilde{x}_{ik}^{T2}(v) = x_k(v)$ for $k \in T$, and the inference (A.3) no longer applies to that variety. Beliefs about unmatched varieties are as under T1. A card showing a variety that combines the two traits can lower the strength of the trade-off view, $\tau_i^{T2} \leq \tau_i^{T1}$.

D. Derivations

Proposition 1 (Knowledge)**.** *In both treatment arms, $\rho_{ik}^{g} > \rho_{ik}^{C}$ for $k \in T$.*

Proof. The session is the same in both arms, and raising the weight on the cues is what it does in the framework. This is the first stage of the experiment, which recall of the cue mapping measures directly, and not a derived result. □

Proposition 2 (Trait identification)**.** *Let $b_{ik}(v) = \tilde{x}_{ik}^{0}(v) - x_k(v)$ denote the error in the prior. (i) Under T2, the belief about each targeted trait of a DNA-matched variety is correct. (ii) Under T1, the expected squared error of the belief falls if and only if*

$$\frac{\rho_{ik}^{C} + \rho_{ik}^{T1}}{2} < \frac{b_{ik}(v)^2}{b_{ik}(v)^2 + \sigma_k^2}. \tag{A.8}$$

Proof. (i) is immediate from $\tilde{x}_{ik}^{T2}(v) = x_k(v)$. (ii) From (1) and (A.2), $\tilde{x}_{ik} - x_k = (1 - \rho_{ik})\, b_{ik} + \rho_{ik}\, \varepsilon_{ik}$, so the expected squared error is $g(\rho) = (1-\rho)^2 b_{ik}^2 + \rho^2 \sigma_k^2$. Then

$$g(\rho_{ik}^{T1}) - g(\rho_{ik}^{C}) = \left(\rho_{ik}^{T1} - \rho_{ik}^{C}\right)\Big[\left(b_{ik}^2 + \sigma_k^2\right)\left(\rho_{ik}^{C} + \rho_{ik}^{T1}\right) - 2\, b_{ik}^2\Big],$$

which is negative if and only if (A.8) holds, since $\rho_{ik}^{T1} > \rho_{ik}^{C}$. □

[26] Exact proportionality in (A.6) is a property of the quadratic form. With any net return that is concave in the area of each variety, the signs and the ranking of the responses are the same.

The right-hand side of (A.8) is the weight that minimizes the error. T1 improves accuracy when it moves the weight toward that value without overshooting it by more than the initial shortfall, and worsens it when the cue is noisy relative to the error in the prior or when farmers come to rely on the rule more than the precision of the cue warrants.

Proposition 3 (Reading the cues and land)**.** *Under T1, for a variety the farmer plants:*

(i) Land follows beliefs: $\Delta^{T1} a_i(v) = \frac{1}{\phi_i} \sum_{k \in T} (\omega_{ik} - \tau_i \, \kappa_{ik}) \, \Delta^{T1} \tilde{x}_{ik}(v)$.

(ii) The belief about vitamin A content moves toward the cue, whatever the true trait:

$$\Delta^{T1} \tilde{x}_{ic}(v) = \Delta\rho_{ic} \left(s_{ic}(v) - \tilde{x}^0_{ic}(v) \right). \tag{A.9}$$

Across a farmer's varieties, the covariance of this change with the true trait is $\Delta\rho_{ic} \operatorname{Cov}\left(x_c - \tilde{x}^0_{ic},\, x_c\right)$*, while its variance is* $\Delta\rho^2_{ic} \left[\operatorname{Var}\left(x_c - \tilde{x}^0_{ic}\right) + \sigma^2_c\right]$*: the noisier the cue, the smaller the part of the reallocation that goes to varieties that carry the trait.*

(iii) The change in perceived value is

$$\Delta^{T1} \tilde{\pi}_i(v) = N_{ic} \, \Delta^{T1} \tilde{x}_{ic}(v) + \left(\omega_{ir} - \tau_i \, \kappa_{ir}\right) \left[\Delta^{T1} \hat{x}_{ir}(v) + \Delta\rho_{ir} \, \tau_i \, \theta \, \tilde{x}^C_{ic}(v)\right], \tag{A.10}$$

$$N_{ic} \equiv \omega_{ic} - \tau_i \, \kappa_{ic} - \tau_i \, \theta \, (1 - \rho^{T1}_{ir}) \left(\omega_{ir} - \tau_i \, \kappa_{ir}\right). \tag{A.11}$$

N_{ic} is the net value of coming to regard a variety as rich in vitamin A: its nutritional value, minus the costs she associates with vitamin A content for other outcomes, minus the resistance she believes she gives up. It is zero or negative when $\tau_i \left[\kappa_{ic} + \theta \, (1 - \rho^{T1}_{ir})(\omega_{ir} - \tau_i \, \kappa_{ir})\right] \geq \omega_{ic}$*, even though* $\omega_{ic} > 0$*. With the inference written in one direction, the corresponding net value of coming to regard a variety as resistant is* $\omega_{ir} - \tau_i \, \kappa_{ir}$*; the remark below gives the case in which it runs both ways.*

Proof. (i) follows from (A.6) and (2) with τ_i and $\eta_i(v)$ unchanged. (ii) Equation (A.9) follows from (1) with the prior and the cue held fixed. Substituting $s_{ic} = x_c + \varepsilon_{ic}$ gives $\Delta\rho_{ic} \left[(x_c - \tilde{x}^0_{ic}) + \varepsilon_{ic}\right]$; the noise is independent of the trait and the prior, which gives the covariance and the variance. (iii) From (A.3),

$$\begin{aligned} \Delta^{T1} \tilde{x}_{ir} &= \Delta^{T1} \hat{x}_{ir} - \tau_i \theta \left[(1 - \rho^{T1}_{ir}) \, \tilde{x}^{T1}_{ic} - (1 - \rho^{C}_{ir}) \, \tilde{x}^{C}_{ic}\right] \\ &= \Delta^{T1} \hat{x}_{ir} - (1 - \rho^{T1}_{ir}) \, \tau_i \theta \, \Delta^{T1} \tilde{x}_{ic} + \Delta\rho_{ir} \, \tau_i \theta \, \tilde{x}^C_{ic}. \end{aligned}$$

Substituting in $\Delta^{T1}\tilde{\pi}_i = \sum_k (\omega_{ik} - \tau_i \kappa_{ik})\, \Delta^{T1}\tilde{x}_{ik}$ and collecting the terms in $\Delta^{T1}\tilde{x}_{ic}$ gives (A.10). The last term in (A.10) is positive: as she relies more on the resistance cue, the inference from flesh color counts for less. □

Proposition 3 says when T1 leaves planting unchanged. With an informative cue and $N_{ic} > 0$, T1 moves land on average toward the varieties that carry the trait. It does not when the net value of the trait is close to zero (part iii), when the cue is noisy enough that the reallocation is mostly unrelated to the true trait (part ii), or when $\Delta\rho_{ik}$ is small. If the session also raises the value farmers place on a trait, a further term proportional to $\tilde{x}^{C}_{ik}(v)$ enters (A.10); it is common to both arms.

The inference in (A.3) is written in one direction for simplicity. Suppose the farmer also expects a variety she regards as resistant to be less rich in vitamin A, so that $\tilde{x}_{ic}(v) = \hat{x}_{ic}(v) - a\,\tilde{x}_{ir}(v)$ and $\tilde{x}_{ir}(v) = \hat{x}_{ir}(v) - b\,\tilde{x}_{ic}(v)$, where $b = (1-\rho_{ir})\,\tau_i\,\theta$ as in (A.3), $a \geq 0$ is its counterpart for vitamin A content, and $ab < 1$. Solving the two equations, the net value of coming to regard a variety as rich in vitamin A is $\big[(\omega_{ic} - \tau_i\kappa_{ic}) - b\,(\omega_{ir} - \tau_i\kappa_{ir})\big]/(1-ab)$, and that of coming to regard it as resistant is $\big[(\omega_{ir} - \tau_i\kappa_{ir}) - a\,(\omega_{ic} - \tau_i\kappa_{ic})\big]/(1-ab)$. When both are positive, each is smaller than the value of the trait taken alone, so part (iii) applies to either trait. Under T2 the card states both traits, so the first expression in the proof of Proposition 4 is unchanged, and the gain from removing the inference has a counterpart for vitamin A content.

Proposition 4 (Reallocation toward varieties with targeted traits)**.** *Let $\Delta_{ic}(v) = x_c(v) - \tilde{x}^{T1}_{ic}(v)$ and $\hat{\Delta}_{ir}(v) = x_r(v) - \hat{x}^{T1}_{ir}(v)$ denote what remains unrecognized after T1 in vitamin A content and, setting aside the inference from flesh color, in resistance. For a DNA-matched variety the farmer plants,*

$$\Delta^{T2}a_i(v) - \Delta^{T1}a_i(v) = \frac{1}{\phi_i}\Big[\underbrace{\big(\omega_{ic} - \tau_i^{T1}\kappa_{ic}\big)\,\Delta_{ic}(v)}_{\textit{vitamin A belief corrected}} + \underbrace{\big(\omega_{ir} - \tau_i^{T1}\kappa_{ir}\big)\,\hat{\Delta}_{ir}(v)}_{\textit{resistance belief corrected}}$$
$$+ \underbrace{\big(\omega_{ir} - \tau_i^{T1}\kappa_{ir}\big)(1-\rho_{ir}^{T1})\,\tau_i^{T1}\,\theta\,\tilde{x}^{T1}_{ic}(v)}_{\textit{inference from flesh color removed}} + \underbrace{\big(\tau_i^{T1} - \tau_i^{T2}\big)\sum_{k\in T}\kappa_{ik}\,x_k(v)}_{\textit{weaker view of trade-offs}}\Big]. \quad \text{(A.12)}$$

Suppose the farmer perceives each trait as valuable on net, $\omega_{ik} > \tau_i^{T1}\kappa_{ik}$. Then:

(i) The first two terms have the sign of what was unrecognized: a variety whose traits were unrecognized ($\Delta_{ic} \geq 0$, $\hat{\Delta}_{ir} \geq 0$, one of them strict) gains land, and the first two terms are negative for a variety that was overvalued.

(ii) If $\tau_i^{T1}\theta > 0$ and $\rho_{ir}^{T1} < 1$, a variety that carries both traits and whose resistance was not overestimated ($\hat{\Delta}_{ir} \geq 0$) gains land whatever the farmer believed about

its vitamin A content, between the reference level and the truth: to the extent that she did not regard it as rich in vitamin A, the first term is positive, and to the extent that she did, the third term is positive.

(iii) For a farmer who already held correct beliefs about both traits of the variety, the first three terms sum to zero and only the last can be positive.

(iv) If the card weakens the view of trade-offs, $\tau_i^{T2} < \tau_i^{T1}$, every matched variety that carries a targeted trait gains through the last term, whether or not its traits were recognized.

Proof. For a matched variety, $\tilde{\pi}_i^{T2} = \bar{\pi} + \sum_k (\omega_{ik} - \tau_i^{T2}\kappa_{ik})\, x_k + \eta_i$ and $\tilde{\pi}_i^{T1} = \bar{\pi} + \sum_k (\omega_{ik} - \tau_i^{T1}\kappa_{ik})\, \tilde{x}_{ik}^{T1} + \eta_i$. Adding and subtracting $\sum_k (\omega_{ik} - \tau_i^{T1}\kappa_{ik})\, x_k$,

$$\tilde{\pi}_i^{T2} - \tilde{\pi}_i^{T1} = \sum_{k \in T} \big(\omega_{ik} - \tau_i^{T1}\kappa_{ik}\big)\big(x_k - \tilde{x}_{ik}^{T1}\big) + \big(\tau_i^{T1} - \tau_i^{T2}\big) \sum_{k \in T} \kappa_{ik}\, x_k.$$

By (A.3), $x_r - \tilde{x}_{ir}^{T1} = \hat{\Delta}_{ir} + (1 - \rho_{ir}^{T1})\, \tau_i^{T1}\theta\, \tilde{x}_{ic}^{T1}$. Substituting and dividing by ϕ_i, using (A.6), gives (A.12). (i) reads off the signs. (ii) With $x_c > 0$ and $0 \le \tilde{x}_{ic}^{T1} \le x_c$, the first and third terms sum to a weighted combination of two positive numbers, $(\omega_{ic} - \tau_i^{T1}\kappa_{ic})\,(x_c - \tilde{x}_{ic}^{T1}) + (\omega_{ir} - \tau_i^{T1}\kappa_{ir})(1 - \rho_{ir}^{T1})\, \tau_i^{T1}\theta\, \tilde{x}_{ic}^{T1}$, which is positive for every value of $\tilde{x}_{ic}^{T1}$ in that range; the second term is non-negative. (iii) If $\tilde{x}_{ik}^{T1} = x_k$ for both traits, the first sum in the display above is zero. (iv) follows from $\kappa_{ik} \ge 0$ and $x_k(v) > 0$. □

Part (ii) is the sense in which a farmer who believes the two traits trade off undervalues a variety that combines them: she cannot regard it as rich in vitamin A without lowering what she believes about its resistance. Parts (ii) and (iii) together give the concentration of the T2 response among farmers who hold such a variety and did not know it. Part (iv) is specific to the second channel: it implies gains on the farmer's other varieties that carry a targeted trait, and fewer perceived weaknesses for the varieties she regards as nutritious or resistant. For varieties that could not be matched, beliefs are as under T1 and only the last channel operates.

Proposition 5 (Total sweet potato area)**.** *Total area changes by*

$$\Delta^{T2} A_i = \frac{1}{\phi_i} \sum_{v \in \mathcal{V}_i} \Delta^{T2} \tilde{\pi}_i(v). \tag{A.13}$$

If, within the season, the farmer acts only on increases in perceived value, total area rises whenever at least one of her varieties gains in perceived value. Once reductions

are also made, total area is higher than in the control group if and only if perceived value rises on net across her varieties.

Proof. Sum (A.6) over the varieties she plants. If only increases are acted on, $\Delta^{T2} A_i = \frac{1}{\phi_i} \sum_v \max\{\Delta^{T2} \tilde{\pi}_i(v), 0\}$, which is positive whenever one term is. □

Whether total area remains higher once reductions are made therefore depends on the balance of upward and downward corrections. In the control group, unrecognized SPVD resistance is more than twice as common as wrongly attributed resistance, while for yellow flesh and AB resistance the two are of similar size (Table 2). Because introducing a variety the farmer does not hold requires (A.7), and the information arrived shortly before planting, these responses are more likely to appear first in the land given to varieties farmers already hold than in the set of varieties they hold.

B Appendix: Intervention Content

This appendix provides the detailed content of the information sessions delivered under both treatment arms of the study. All sessions were conducted by trained facilitators using visual aids and printed materials, ensuring consistent delivery across villages.

A. General Information Session (Treatment Arms 1 and 2)

Topic 1: Nutritional Content and Sweet Potato Flesh Color

Facilitators explained the nutritional differences between sweet potato varieties based on flesh color using visual aids. The following key messages were delivered:

- Sweet potato varieties have different flesh colors: orange, yellow, white, and purple.
- The flesh color indicates the vitamin A content of the variety:
  - Orange-fleshed varieties have high vitamin A content.
  - Yellow-fleshed varieties contain some vitamin A, but less than orange varieties.
  - White and purple-fleshed varieties contain little or no vitamin A.
- Vitamin A is important for:
  - Maintaining a healthy immune system.
  - Preventing night blindness and age-related vision decline.
  - Supporting healthy growth, particularly in children and pregnant women.

Topic 2: Disease Identification and Management

Facilitators used visual aids to show symptoms of two common sweet potato diseases: Sweet Potato Virus Disease (SPVD) and Alternaria Blight (AB). The following key messages were conveyed:

- Diseased plants become weak and are more susceptible to further infection.
- Pests and insects spread these diseases, similar to how mosquitoes transmit malaria to humans.
- While pesticides may kill insects that spread the disease, they do not cure already infected plants.

- Farmers can help prevent disease spread by:
  - Uprooting and removing diseased plants promptly.
  - Avoiding the use of infected vines for planting in the next season.

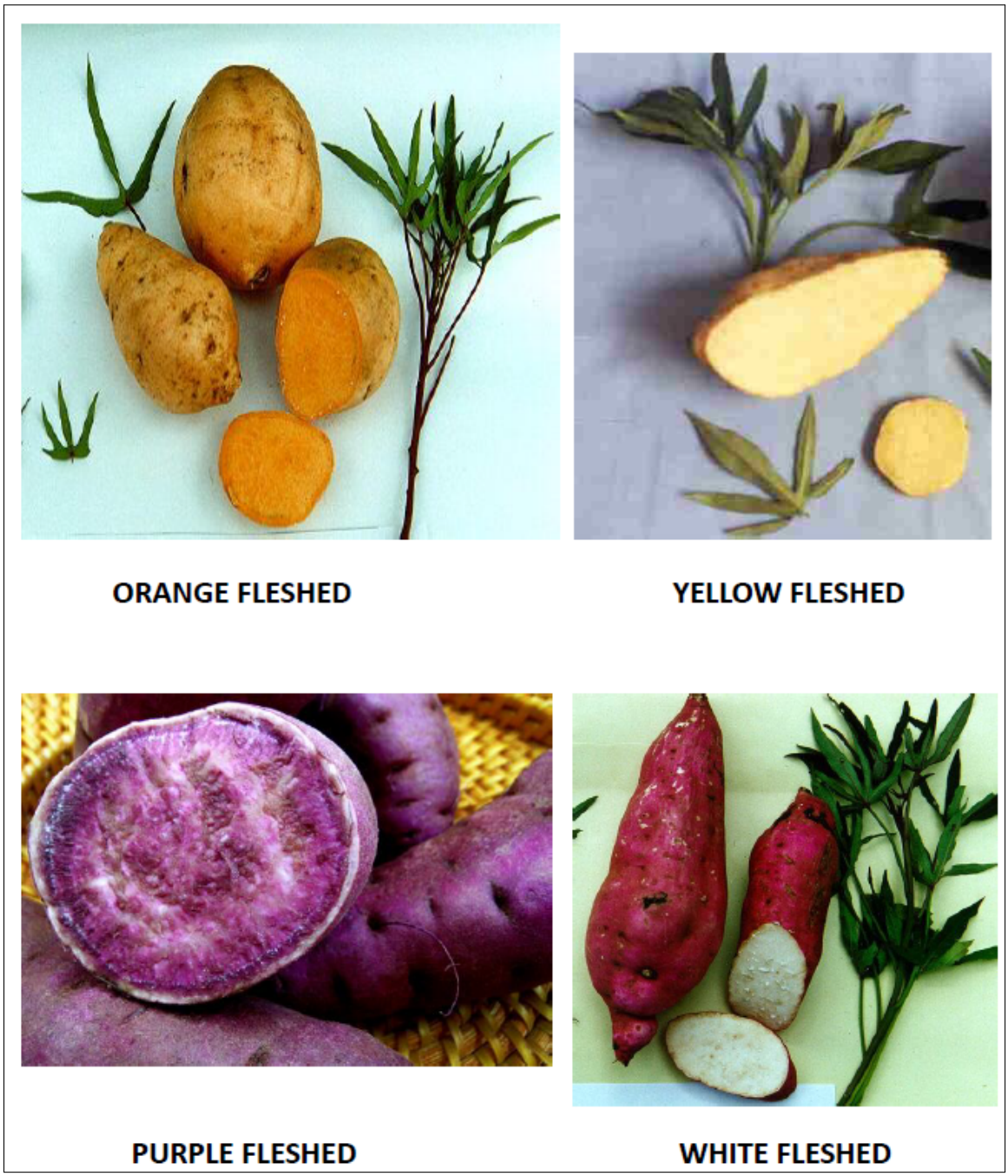


Figure A1: Visual aid used to illustrate different sweet potato flesh colors

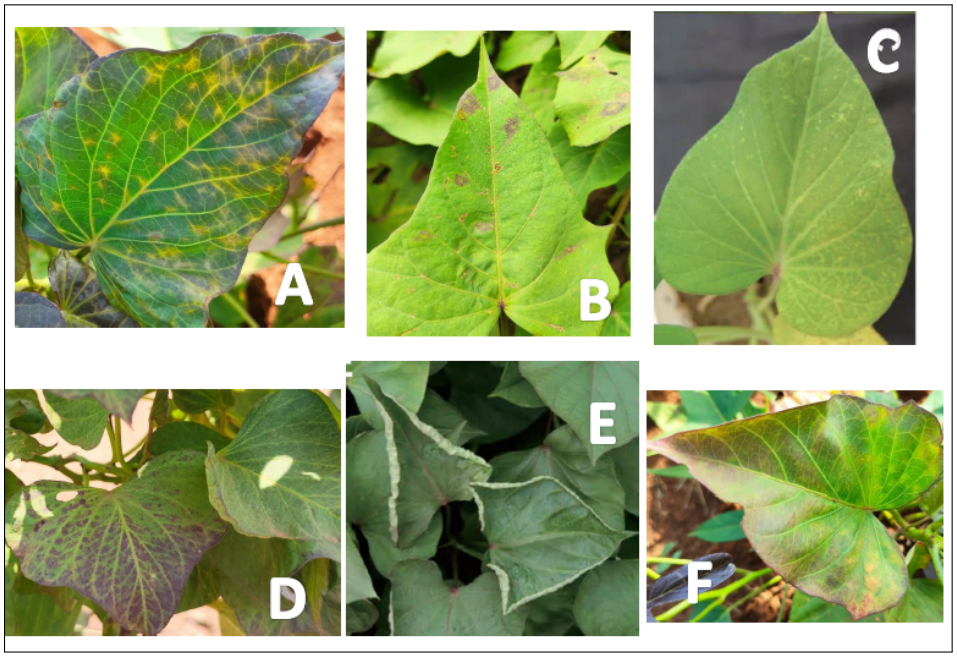


Figure A2: Visual aid used to illustrate SPVD symptoms

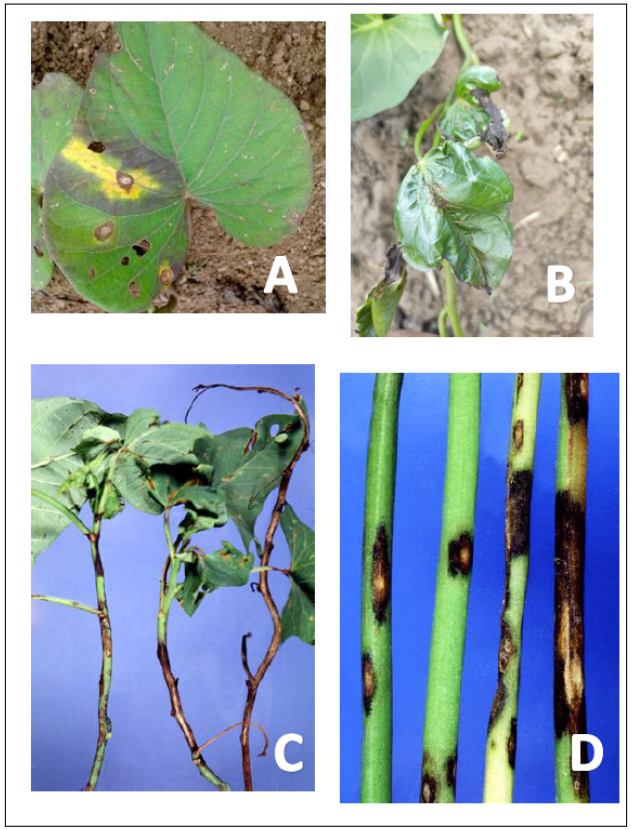


Figure A3: Visual aid used to illustrate AB symptoms

B. Variety-Specific Trait Information: Individual and Community Cards (Treatment Arm 2 Only)

In villages assigned to Treatment Arm 2, farmers received additional information based on DNA fingerprinting results.

Group 1: Individual Feedback Cards

Farmers whose plots were sampled received a printed individual card containing:

- Farmer's name, district, parish, subcounty, and village.
- A list of varieties grown by the farmer, as identified through DNA fingerprinting.
- The official names of the varieties (if available).
- Trait information for each variety:
  - Resistance to SPVD.
  - Resistance to Alternaria Blight.
  - Overall disease resistance.
  - Flesh color (linked to vitamin A content).
- Color photographs illustrating healthy plants and disease symptoms for easier identification.

Group 2: Community-Level Summary Cards

Farmers who did not participate in DNA sampling received a community card aggregating varietal and trait information from sampled farmers in their community. This card:

- Listed the varieties most commonly found in the village.
- Presented aggregated information on disease resistance and nutritional content.
- Was used as a discussion tool during the session and provided as a take-home reference.

Anonymized copies of the individual and community cards used in the intervention are available upon request and illustrated in Figure A4.

District	Parish	Subcounty	Village	Name
A	B	C	D	E





Variety	Official name	Resistant to SPVD	Resistant to AB	Disease resistant	Flesh color
Funtula	Muwulu Aduduma	YES	NO	NO	White or purple
Icok agamente	KAKAMEGA	YES	YES	YES	Orange
Lira lira	KML934	YES	YES	YES	Yellow

Figure A4: Example of individual summary card

C Appendix: Tables

Table A1: Composition of Reference Library and Sample Matching Outcomes

	Original Reference Library	With Cluster-Based Imputation
Total number of varieties	133	133
Varieties with recorded flesh color	86	–
Varieties classified by orange flesh color	–	133
Varieties classified by yellow flesh color	–	102
Varieties with SPVD resistance information	90	109
Varieties with Alternaria blight (AB) resistance information	90	109
Field Samples Matched to Reference Library		
Total number of farmer samples collected	3,316	
Number of samples successfully matched	1,641 (49%)	
Matched samples with orange flesh classification	1,641	
Matched samples with yellow flesh classification	1,210	
Matched samples with virus resistance classification	1,264	

Table A2: Intent-to-Treat Effects on Recall of Sweet Potato Information, by Respondent Gender.

	(1) Color→ Nutrition	(2) Orange = Best	(3) SPVD: Symptoms	(4) SPVD: Control	(5) AB: Symptoms	(6) AB: Control
T1	0.117***	0.0700*	1.247***	0.645***	0.634**	0.360*
	(0.043)	(0.040)	(0.287)	(0.174)	(0.269)	(0.188)
T1*Woman	0.0700	0.0784*	-0.269	0.179	0.251	0.382*
	(0.050)	(0.046)	(0.356)	(0.197)	(0.283)	(0.203)
T2	0.0866**	0.0899**	0.749**	0.600***	0.833***	0.386*
	(0.040)	(0.045)	(0.331)	(0.181)	(0.255)	(0.208)
T2*Woman	0.114**	0.0869	-0.0424	0.200	0.0000907	0.214
	(0.051)	(0.057)	(0.421)	(0.217)	(0.296)	(0.226)
Fem. Resp.	-0.0566	-0.0644**	0.0951	-0.146	-0.00336	-0.280*
	(0.037)	(0.031)	(0.270)	(0.141)	(0.216)	(0.159)
Obs.	2105	2105	2105	2105	2105	2105
Control Mean	0.62	0.23	-0.08	0.10	-0.00	0.20
p-value: T1=T2 (women)	0.63	0.35	0.19	0.88	0.79	0.22

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on respondents' recall of key messages, allowing effects to differ by the randomly assigned respondent's gender. The unit of observation is the respondent. "Fem. Resp." equals 1 if the assigned respondent was a woman, alone or jointly with a man, and 0 if a man alone; "T1×Woman" and "T2×Woman" are the corresponding interactions, so the T1 and T2 rows give the effects for male respondents and the interactions give the female–male difference. Outcomes are as in Table 3. All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. The "Control Mean" row reports the mean among control-group households with a male respondent. The last row reports the p-value from a test of equality of the T1 and T2 effects for female respondents. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Differential Attrition, Sweet Potato Cultivation, DNA-Fingerprinting Participation, and Program Take-Up by Treatment Assignment.

	(1) Pr(Attrition at Endline)	(2) Pr(HH Cultivated SP in 2nd Season)	(3) Pr(Attrition in DNA-Fingerprinting)	(4) Pr(Attrition Health Module)	(5) Pr(Attrition Consumption Module)	(6) Pr(Take-Up)
T1	0.00155	0.00151	0.0338	0.00584	0.0120	-0.0200
	(0.01)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)
T2	-0.00674	0.00623	0.0383	0.0102	0.0149	—
	(0.01)	(0.02)	(0.03)	(0.01)	(0.01)	
Observations	2150	2105	1012	5421	7187	1428
Mean Control or T2 Mean	0.02	0.84	0.18	0.10	0.15	0.93
P-value T1=T2	0.27	0.81	0.88	0.67	0.79	0.15

Notes: Each column reports coefficients from a separate OLS regression of the indicator shown in the column heading on treatment assignment indicators. Coefficients on T1 and T2 are differences relative to the pure control group in columns (1)–(5). Column (6) reports differences in program take-up between T1 and the T2 group (which is the omitted category), hence the T2 coefficient is not shown and the row "Mean Control or T2 Mean" reports the mean take-up rate in T2. "Pr(Attrition at Endline)" equals 1 if the household did not complete the endline survey. "Pr(HH Cultivated SP in 2nd Season)" equals 1 if the household cultivated sweet potato in the second season at endline. "Pr(Attrition in DNA-Fingerprinting)" equals 1 if the household did not consent to leaf collection, attrited prior to DNA collection, or did not cultivate sweet potato in the relevant season. "Pr(Attrition Health Module)" and "Pr(Attrition Consumption Module)" equal 1 if the baseline reference child was not living in the household and thus the corresponding module could not be completed. All specifications include stratum fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Baseline Characteristics by Treatment Arm.

	Means			P-values			
	Control (C)	T1	T2	C=T1	C=T2	T1=T2	N
A. Household characteristics and information							
Female hh-head	.214	.240	.238	.176	.292	.920	2141
HH-head education (years)	1.914	1.922	1.908	.976	.750	.908	2141
HH member in off-farm employment	.614	.626	.632	.634	.564	.902	2141
Household size	6.362	6.894	6.554	.002	.064	.050	2141
Dependency ratio	1.152	1.196	1.148	.584	.740	.576	2141
At least one pregnant woman	.224	.212	.226	.486	.762	.458	2141
Number of children below age 5	.708	.678	.730	.470	.350	.202	2141
HH has heard about Vitamin A	.642	.654	.654	.910	.812	.926	2141
Extension agent visit in the last 12 months	.482	.446	.552	.632	.420	.358	2141
Cultivated SP in season 2	.786	.740	.752	.146	.490	.564	2141
B. Sweet Potato Production							
Total cultivated area (m^2) (GPS)	480.09	514.24	475.00	0.898	0.948	0.990	845
Quantity harvested (kg)	86.82	82.78	116.46	0.506	0.190	0.176	1012
Quantity consumed (kg)	108.75	75.69	121.42	0.258	0.700	0.232	1012
C.Adoption of at least one variety							
Orange-fleshed	0.054	0.062	0.040	0.754	0.710	0.326	1012
Yellow-fleshed	0.232	0.224	0.278	0.850	0.120	0.310	1012
SPVD resistant	0.530	0.526	0.506	0.582	0.458	0.620	1012
Alternaria blight resistant	0.444	0.446	0.468	0.886	0.476	0.626	1012
D. Share of varieties grouped by traits (DNA-Fingerprinting)							
Orange flesh	.018	.024	.020	.430	.616	.864	1000
Yellow flesh	.124	.132	.166	.210	.016	.428	1000
SPVD Resistant	.296	.306	.320	.876	.348	.302	1000
AB Resistant	.242	.228	.280	.292	.174	.030	1000
Not identified	.508	.508	.510	.332	.642	.708	1000
E.Child Diet - Age under 5							
Diet diversity score (child)	8.516	8.518	8.416	0.802	0.894	0.830	507
Child ate sweet potato in last 7 days	0.736	0.630	0.616	0.010	0.030	0.880	505
Child ate orange/yellow SP in last 7 days	0.166	0.200	0.238	0.804	0.216	0.328	505

Notes: This table reports baseline means of household, production, varietal adoption, DNA-fingerprinting, and child diet variables by experimental assignment (control, T1, and T2). The final three columns report p-values from two-sided tests of equality of means for the indicated pairwise comparisons within each row. The overall baseline sample includes 2,141 households. Variables based on DNA fingerprinting are measured for the subsample of 1,000 households that provided leaf samples. Sweet potato production outcomes are measured for the subsamples indicated by the N column. Child diet outcomes are measured for the sample of children under age 5 shown in Panel E (N varies by outcome).

Table A5: Intent-to-Treat Effects on Household Area Cultivated by Variety Traits, Without Winsorization (Square Meters).

	(1) Orange	(2) Yellow	(3) SPVD	(4) AB	(5) Yellow & Virus res.	(6) White	(7) Purple	(8) Not identified	(9) GPS total area (sqm)
T1	-5.081	-16.24	24.97	4.770	-5.943	41.66	1.019	6.577	0.493
	(4.60)	(22.28)	(34.70)	(27.18)	(22.35)	(25.80)	(1.21)	(22.28)	(44.18)
T2	-1.086	58.11*	75.00*	91.91**	68.49**	23.60	1.460	14.86	101.1*
	(4.56)	(30.36)	(41.46)	(37.95)	(30.93)	(27.86)	(1.34)	(25.19)	(51.84)
Observations	1012	1012	1012	1012	1012	1012	1012	1012	1012
Mean Control	11.47	91.78	199.03	145.95	64.65	103.40	0.29	178.59	552.38
P-value T1=T2	0.30	0.06	0.30	0.04	0.07	0.57	0.82	0.76	0.05

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on household land area planted with sweet potato varieties grouped by DNA-identified traits. Each column corresponds to the total area (in square meters) cultivated with varieties possessing the trait indicated in the column heading; column (9) reports total sweet-potato area measured by GPS. For columns (1)–(8), the area planted to each trait category is constructed by multiplying GPS-measured plot area by the farmer's self-reported percentage of the plot allocated to each variety, and then aggregating across plots within the household. All specifications are estimated by OLS and include stratum fixed effects and baseline covariates selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Intent-to-Treat Effects on the Number of Sweet Potato Varieties Planted by Trait Category (Extensive Margin).

	(1) Orange	(2) Yellow	(3) SPVD	(4) AB	(5) SPVD/AB	(6) Orange/Yellow and Virus res.	(7) No trait	(8) Total
T1	-0.0131	0.0797	0.109	0.0845	0.104	0.0403	-0.161	0.0000499
	(0.02)	(0.07)	(0.09)	(0.07)	(0.09)	(0.06)	(0.13)	(0.14)
T2	0.0107	0.0308	0.122	0.198***	0.134	0.0906	0.114	0.296*
	(0.02)	(0.06)	(0.09)	(0.08)	(0.09)	(0.06)	(0.14)	(0.18)
Observations	1012	1012	1012	1012	1012	1012	1012	1012
Mean Control	0.06	0.49	1.11	0.88	1.12	0.39	2.20	3.67
P-value T1=T2	0.30	0.54	0.88	0.17	0.74	0.47	0.09	0.10

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on the number of sweet potato varieties planted by trait category. The unit of observation is the household. In each column, the dependent variable is the number of varieties cultivated by the household that fall into the category shown in the column heading. Trait categories are based on DNA-verified varietal traits. "SPVD" and "AB" refer to resistance to sweet potato virus disease and Alternaria blight, respectively; "SPVD/AB" indicates varieties resistant to either SPVD or AB; "Orange/Yellow and Virus res." indicates varieties that are orange- or yellow-fleshed and virus-resistant; and "No trait" indicates varieties not classified into the listed trait categories. All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. The "Mean Control" row reports the control-group mean of the dependent variable. The row "p-value T1=T2" reports the p-value from a test of equality of the treatment coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Intent-to-Treat Effects on Adoption of at Least One Sweet Potato Variety by Trait Category.

	(1) Orange	(2) Yellow	(3) SPVD	(4) AB	(5) SPVD/AB	(6) Orange/Yellow and Virus res.
T1	-0.0199	0.0119	0.0536	0.0302	0.0498	0.00638
	(0.02)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)
T2	0.0110	0.00324	0.00248	0.0331	-0.00110	0.0367
	(0.02)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
Observations	1012	1012	1012	1012	1012	1012
Mean Control	0.06	0.40	0.73	0.65	0.73	0.34
P-value T1=T2	0.12	0.87	0.14	0.94	0.14	0.57

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on adoption of at least one sweet potato variety by trait category. The unit of observation is the household. In each column, the dependent variable is an indicator equal to 1 if the household cultivates at least one variety in the category shown in the column heading, and 0 otherwise. Trait categories are based on DNA-verified varietal traits. "SPVD" and "AB" refer to resistance to sweet potato virus disease and Alternaria blight, respectively; "SPVD/AB" indicates varieties resistant to either SPVD or AB; and "Orange/Yellow and Virus res." indicates varieties that are orange- or yellow-fleshed and virus-resistant. All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. The "Mean Control" row reports the control-group mean of the dependent variable. The row "p-value T1=T2" reports the p-value from a test of equality of the treatment coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Intent-to-Treat Effects on Household Sweet Potato Yield (Kilograms per Square Meter).

	(1) Yields (kg/sqm)
T1	-0.0624
	(0.05)
T2	-0.0842
	(0.05)
Obs.	1010
Control Mean	0.36
P-value T1=T2	0.28

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on household sweet potato yield, measured as total production (kilograms) divided by total cultivated area (square meters). The unit of observation is the household. The dependent variable is yield in kilograms per square meter. All specifications are estimated by OLS and include stratum fixed effects; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. The "Control Mean" row reports the control-group mean of the dependent variable. The row "p-value T1=T2" reports the p-value from a test of equality of the treatment coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Intent-to-Treat Effects on Consumption of Orange- or Yellow-Fleshed Sweet Potato among Children under Age 5.

	Last 7 days		Last 6 months
	(1) Any Consumption	(2) Number of times consumed	(3) Consumed Once a week
T1	0.00739	-0.281	0.00459
	(0.05)	(0.33)	(0.05)
T2	0.0221	-0.0597	0.0218
	(0.05)	(0.35)	(0.04)
Obs.	681	681	681
Control Mean	0.55	2.80	0.64
p-value: T1=0	0.887	0.398	0.919
RI p-value: T1=0	0.877	0.479	0.933
p-value: T2=0	0.649	0.865	0.625
RI p-value: T2=0	0.706	0.878	0.645
p-value: T1=T2	0.777	0.479	0.706
RI p-value: Joint	0.930	0.760	0.887

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on consumption of orange- or yellow-fleshed sweet potato among children under age 5. The unit of observation is the child. Columns (1) and (2) use outcomes measured over the last 7 days: "Any Consumption" is an indicator equal to 1 if the child consumed orange- or yellow-fleshed sweet potato at least once, and "Number of times consumed" is the reported number of consumption occasions. Column (3) uses outcomes measured over the last 6 months: "Consumed Once a week" is an indicator equal to 1 if the child consumed orange- or yellow-fleshed sweet potato at least once per week on average. All specifications are estimated by OLS and include stratum fixed effects, child age fixed effects, and a child gender indicator; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. Rows labeled "p-value" report two-sided p-values from conventional inference; rows labeled "RI p-value" report randomization-inference p-values based on 1,000 iterations using the randomization-c procedure. The row "RI p-value: Joint" reports the randomization-inference p-value for a joint test that both treatment coefficients are equal to zero. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Intent-to-Treat Effects on Food Group Consumption in the Last 24 Hours among Children under Age 2.

	(1) Cereals	(2) Roots and tubers	(3) Vegetables	(4) Fruit	(5) Meat	(6) Eggs	(7) Fish	(8) Milk	(9) Legumes	(10) Oils	(11) Sugar	(12) Spices	(13) Food diversity index
T1	-0.0250	-0.0273	-0.0701	0.0109	0.0351	-0.00134	0.0175	-0.0426	-0.000943	-0.00774	-0.0225	0.0000891	-0.407
	(0.04)	(0.08)	(0.05)	(0.07)	(0.07)	(0.05)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)	(0.07)	(0.34)
T2	-0.00846	0.0905	0.0236	0.167**	0.0639	0.0166	0.0519	0.0224	0.0885	-0.0333	0.0794	-0.0430	0.465
	(0.05)	(0.08)	(0.04)	(0.07)	(0.07)	(0.06)	(0.08)	(0.09)	(0.08)	(0.08)	(0.06)	(0.06)	(0.32)
Obs.	202	202	202	202	202	202	202	202	202	202	202	202	202
Control Mean	0.91	0.71	0.91	0.47	0.17	0.12	0.29	0.39	0.77	0.65	0.75	0.77	6.89
p-value: T1=0	0.529	0.724	0.167	0.883	0.603	0.978	0.815	0.557	0.989	0.909	0.629	0.999	0.226
RI p-value: T1=0	0.642	0.766	0.207	0.885	0.647	0.983	0.811	0.614	0.993	0.924	0.723	0.999	0.313
p-value: T2=0	0.865	0.241	0.527	0.013	0.327	0.772	0.517	0.804	0.244	0.670	0.187	0.473	0.144
RI p-value: T2=0	0.891	0.296	0.691	0.041	0.360	0.777	0.509	0.815	0.263	0.685	0.191	0.549	0.226
p-value: T1=T2	0.768	0.201	0.106	0.055	0.680	0.770	0.712	0.468	0.257	0.777	0.102	0.506	0.026
RI p-value: Joint	0.896	0.383	0.234	0.087	0.670	0.947	0.808	0.774	0.431	0.912	0.221	0.806	0.111

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on child food consumption in the last 24 hours among children under age 2. The unit of observation is the child. In columns (1)–(12), each dependent variable is an indicator equal to 1 if the caregiver reports that the child consumed at least one item from the food group listed in the column heading during the previous 24 hours. Column (13) reports a food diversity index defined as the count of food groups (among columns (1)–(12)) consumed in the previous 24 hours. All specifications are estimated by OLS and include stratum fixed effects, child age fixed effects, and a child gender indicator; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. Rows labeled "p-value" report two-sided p-values from conventional inference; rows labeled "RI p-value" report randomization-inference p-values based on 1,000 iterations using the randomization-c procedure. The row "RI p-value: Joint" reports the randomization-inference p-value for a joint test that both treatment coefficients are equal to zero. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Intent-to-Treat Effects on Diarrhea Outcomes among Children under Age 5.

	(1) Last 2 weeks	(2) Days last episode	(3) Last 6 months
T1	-0.0453	-0.298*	-0.0661
	(0.03)	(0.17)	(0.04)
T2	-0.0516	-0.370*	-0.0916*
	(0.03)	(0.20)	(0.05)
Obs.	923	923	923
Control Mean	0.17	1.52	0.42
p-value: T1=0	0.148	0.084	0.125
RI p-value: T1=0	0.259	0.143	0.190
p-value: T2=0	0.115	0.060	0.052
RI p-value: T2=0	0.186	0.090	0.069
RI p-value: Joint	0.376	0.173	0.175

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on diarrhea outcomes among children under age 5. The unit of observation is the child. Column (1) uses an indicator equal to 1 if the child had diarrhea in the last 2 weeks. Column (2) reports the number of days in the child's most recent diarrhea episode. Column (3) uses an indicator equal to 1 if the child had diarrhea at any point in the last 6 months. All specifications are estimated by OLS and include stratum fixed effects, child age fixed effects, and a child gender indicator; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. Rows labeled "p-value" report two-sided p-values from conventional inference; rows labeled "RI p-value" report randomization-inference p-values based on 1,000 iterations using the randomization-c procedure. The row "RI p-value: Joint" reports the randomization-inference p-value for a joint test that both treatment coefficients are equal to zero. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Intent-to-Treat Effects on Diarrhea Outcomes among Children Aged 2–5.

	(1) Last 2 weeks	(2) Days last episode	(3) Last 6 months
T1	-0.0264	-0.135	-0.0482
	(0.04)	(0.15)	(0.04)
T2	-0.0155	-0.0688	-0.0544
	(0.03)	(0.17)	(0.04)
Obs.	707	707	707
Control Mean	0.12	1.06	0.35
p-value: T1=0	0.452	0.365	0.277
RI p-value: T1=0	0.488	0.427	0.307
p-value: T2=0	0.623	0.686	0.196
RI p-value: T2=0	0.686	0.699	0.248
RI p-value: Joint	0.781	0.730	0.441

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on diarrhea outcomes among children aged 2–5. The unit of observation is the child. Column (1) uses an indicator equal to 1 if the child had diarrhea in the last 2 weeks. Column (2) reports the number of days in the child's most recent diarrhea episode. Column (3) uses an indicator equal to 1 if the child had diarrhea at any point in the last 6 months. All specifications are estimated by OLS and include stratum fixed effects, child age fixed effects, and a child gender indicator; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. Rows labeled "p-value" report two-sided p-values from conventional inference; rows labeled "RI p-value" report randomization-inference p-values based on 1,000 iterations using the randomization-c procedure. The row "RI p-value: Joint" reports the randomization-inference p-value for a joint test that both treatment coefficients are equal to zero. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: Spillover Effects of Treatment 2 on Orange/Yellow Sweet Potato Consumption among Children under Age 2 in Community-Information Households, by Tercile of Baseline Village Varietal Shares.

	Last 7 days		Last 6 months
	(1)	(2)	(3)
	Incidence	Nb. of times	At Least Once/Week
T2 - 1st tercile	0.0639	-0.506	0.190
	(0.147)	(0.541)	(0.125)
T2 - 2nd tercile	0.356	1.070	0.0481
	(0.237)	(1.249)	(0.141)
T2 - 3rd tercile	0.232	-0.248	0.160
	(0.196)	(0.598)	(0.195)
Observations	213	213	213
Mean Control	0.46	2.16	0.54
P-value T2(1st)+T2(2nd)+T2(3rd)=0	0.06	0.84	0.16
P-value T2(1st)=T2(2nd)	0.29	0.24	0.42
P-value T2(1st)=T2(3rd)	0.48	0.71	0.90
P-value T2(2nd)=T2(3rd)	0.68	0.32	0.63

Notes: This table reports spillover (indirect) effects of Treatment 2 (T2) on consumption of orange- or yellow-fleshed sweet potato among children under age 2 in households that did not receive the intervention directly but received community cards distributed in T2 villages. These cards report DNA-fingerprinting information on varietal traits of sweet potato varieties cultivated by individual-information households in the community. The unit of observation is the child. Columns (1) and (2) use outcomes measured over the last 7 days: "Incidence" is an indicator equal to 1 if the child consumed orange- or yellow-fleshed sweet potato at least once, and "Nb. of times" is the reported number of consumption occasions. Column (3) uses outcomes measured over the last 6 months: "At Least Once/Week" is an indicator equal to 1 if the child consumed orange- or yellow-fleshed sweet potato at least once per week on average. Coefficients labeled "T2 – 1st tercile" through "T2 – 3rd tercile" report effects estimated separately for community-information households residing in villages in the first, second, and third terciles of the baseline distribution of village varietal shares (treatment intensity). The comparison group consists of children in the pure control group and in T1 villages (which did not receive community cards). All specifications are estimated by OLS and include stratum fixed effects, child age fixed effects, and a child gender indicator; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. Rows labeled "p-value" report two-sided p-values from tests of joint significance across terciles and equality of coefficients across terciles, as indicated. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A14: Spillover Effects of Treatment 2 on Diarrhea Outcomes among Children under Age 5 in Community-Information Households.

	(1) Last 2 weeks	(2) Days last episode	(3) Last 6 months
T2 (G2)	-0.0462	0.164	-0.0259
	(0.03)	(0.27)	(0.04)
Obs.	954	954	954
Control Mean	0.16	1.37	0.40

Notes: This table reports spillover (indirect) effects of Treatment 2 (T2) on diarrhea outcomes among children under age 5 in households that did not receive the intervention directly but received community cards distributed in T2 villages. These cards report DNA-fingerprinting information on varietal traits of sweet potato varieties cultivated by individual-information households in the community. The unit of observation is the child. Column (1) uses an indicator equal to 1 if the child had diarrhea in the last 2 weeks. Column (2) reports the number of days in the child's most recent diarrhea episode. Column (3) uses an indicator equal to 1 if the child had diarrhea at any point in the last 6 months. The coefficient "T2 (G2)" compares community-information households in T2 villages to the comparison group consisting of children in the pure control group and in T1 villages (which did not receive community cards). All specifications are estimated by OLS and include stratum fixed effects, child age fixed effects, and a child gender indicator; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A15: Spillover Effects of Treatment 2 on Diarrhea Outcomes among Children under Age 2 in Community-Information Households, by Tercile of Baseline Village Varietal Shares.

	(1) Last 2 weeks	(2) Days last episode	(3) Last 6 months
T2 - 1st tercile	-0.169	-0.811*	-0.0579
	(0.112)	(0.448)	(0.111)
T2 - 2nd tercile	-0.0475	-0.596	-0.0923
	(0.240)	(1.958)	(0.298)
T2 - 3rd tercile	-0.288***	0.0915	0.161
	(0.097)	(0.990)	(0.174)
Observations	226	226	226
Mean Control	0.31	2.45	0.60
P-value T2(1st)+T2(2nd)+T2(3rd)=0	0.09	0.58	0.98
P-value T2(1st)=T2(2nd)	0.64	0.91	0.91
P-value T2(1st)=T2(3rd)	0.40	0.37	0.27
P-value T2(2nd)=T2(3rd)	0.35	0.75	0.45

Notes: This table reports spillover (indirect) effects of Treatment 2 (T2) on diarrhea outcomes among children under age 2 in households that did not receive the intervention directly but received community cards distributed in T2 villages. These cards report DNA-fingerprinting information on varietal traits of sweet potato varieties cultivated by individual-information households in the community. The unit of observation is the child. Column (1) uses an indicator equal to 1 if the child had diarrhea in the last 2 weeks. Column (2) reports the number of days in the child's most recent diarrhea episode. Column (3) uses an indicator equal to 1 if the child had diarrhea at any point in the last 6 months. Coefficients labeled "T2 – 1st tercile" through "T2 – 3rd tercile" report effects estimated for community-information households residing in villages in the first, second, and third terciles of the baseline distribution of village varietal shares (treatment intensity). The comparison group consists of children in the pure control group and in T1 villages (which did not receive community cards). All specifications are estimated by OLS and include stratum fixed effects, child age fixed effects, and a child gender indicator; additional baseline controls are selected using the post-double-selection LASSO procedure. Standard errors are clustered at the village level and reported in parentheses. Rows labeled "p-value" report two-sided p-values from tests of joint significance across terciles and equality of coefficients across terciles, as indicated. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A16: Intent-to-Treat Effects on Household Land Allocation among Households with a Baseline Win-Win Variety, by Share of Win-Win Varieties Not Recognized as Such (Square Meters).

	(1) Orange	(2) Yellow	(3) SPVD	(4) AB	(5) Yellow & Virus res.	(6) White	(7) Purple	(8) Not identified	(9) GPS total area (sqm)
T1	0.330	-152.2**	-111.1	-124.7*	-132.8**	39.14	-0.629	-57.90	-17.55
	(15.99)	(74.41)	(69.92)	(66.90)	(67.11)	(41.02)	(0.54)	(97.54)	(147.96)
T1 × Share (FN)	-5.464	177.8**	98.22	159.4*	126.2*	-59.03	0.611	27.91	94.30
	(19.93)	(80.95)	(101.37)	(89.21)	(69.60)	(47.26)	(0.82)	(105.34)	(206.33)
T2	16.26	9.276	76.71	24.98	60.36	65.77	-0.342	-18.86	12.65
	(16.95)	(51.98)	(89.27)	(88.13)	(52.22)	(72.73)	(0.67)	(70.91)	(196.37)
T2 × Share (FN)	-19.96	119.1*	100.2	186.0	35.02	6.908	2.202	-9.887	283.6
	(20.60)	(67.19)	(127.06)	(116.95)	(54.24)	(86.16)	(1.42)	(70.50)	(239.89)
Share (FN)	-5.746	-100.3*	-107.5	-156.2**	-68.58	30.67	-1.335	-70.78	-114.8
	(18.70)	(56.79)	(91.64)	(74.06)	(49.83)	(48.31)	(0.94)	(68.68)	(171.33)
Observations	224	224	224	224	224	224	224	224	224
Control mean	20.01	193.78	286.42	209.47	160.19	103.87	0.17	215.36	611.27
p-value: T1 + T1 × Share (FN) = 0	0.62	0.64	0.86	0.59	0.91	0.60	0.98	0.53	0.57
p-value: T2 + T2 × Share (FN) = 0	0.68	0.02	0.04	0.00	0.05	0.15	0.22	0.53	0.01

Notes: This table reports intent-to-treat estimates of the effects of the information treatments on household land allocation (in square meters) to sweet potato varieties grouped by DNA-verified traits, among DNA-sampled households that had at least one win-win variety at baseline. A variety is win-win if DNA fingerprinting identifies it as orange- or yellow-fleshed and resistant to SPVD or AB. Share (FN) is the share of the household's baseline win-win varieties that the farmer did not report as both orange- or yellow-fleshed and SPVD-resistant (false negatives); the baseline survey did not elicit self-reported AB resistance. The unit of observation is the household. In columns (1)–(8), the dependent variable is the total household area cultivated with varieties in the category shown in the column heading; in column (9), it is total sweet potato area measured by GPS. T1 and T2 report the treatment effects for households that recognized all of their win-win varieties (Share (FN) = 0); the interaction terms report how the effects change as Share (FN) rises to one. The last two rows report p-values for the treatment effects among households that recognized none of their win-win varieties (Share (FN) = 1). Control mean is the mean among control households in the estimation sample. All regressions include strata fixed effects and baseline covariates selected by post-double-selection LASSO. Standard errors clustered at the village level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Appendix: Full List of Baseline Controls Used in LASSO Post-Double Selection

This appendix lists the full set of baseline covariates included as candidates in the post-double selection LASSO procedure used to select controls for all regression specifications. Variables were drawn from several domains: demographic characteristics, agricultural practices, sweet potato production history, nutrition and health knowledge, extension exposure, and access to services. Table A17 categorizes and summarizes these variables.

Table A17: Candidate Variables for LASSO Post-Double Selection

Domain	Variables
Demographics	Age, gender of household head, household size, education (years and levels), dependency ratio, off-farm employment, female-headed household indicator, pregnancy status
Nutrition Knowledge & Beliefs	Knowledge about Vitamin A (sources, symptoms, effects), beliefs about sweet potato benefits (appetite, energy, eyesight, immunity, etc.), "don't know" responses
Agricultural Production	Land area, plot counts, yield, tuber size, labor input (paid/unpaid by task), bean/sweet potato production in last 5 years, drought exposure by year
Sweet Potato Specifics	Area, number, and share of varieties (orange, yellow, SPVD-resistant), adoption history, preference for traits (resistance, yield, taste, nutrition, etc.)
Diet and Consumption	Sweet potato consumption (by women and children), frequency, share of varieties consumed, food diversity indices, consumption 24h recall.
Extension and Information	Contact with extension officers, number of visits, information source variables (radio, TV, print, NGO, church, etc.), nutrition or crop-specific training
Access and Infrastructure	Distance to market, town, services, vendor/supplier counts
Health Variables	General and common health indicators, health beliefs (child and adult)
NGO and Program Exposure	Access to credit, seeds, feeding programs, pest management, marketing support, distribution of inputs, year-specific interventions

Note: Final models include only variables selected in either the outcome or treatment

LASSO steps. The complete candidate set ensures robustness of variable inclusion without overfitting.

E Appendix: Bounding the Role of Misclassification in Consumption Outcomes

A. Analytical Framework

Self-reported consumption data may be subject to misclassification errors when respondents cannot accurately identify varietal traits. To assess the robustness of my results to this potential bias, I conduct a bounding analysis following Lewbel (2007) and Kreider and Pepper (2007). This analysis quantifies how much misclassification would be required to fully explain the observed treatment effects without any true behavioral change. I decompose the observed treatment effect on reported consumption into two components: true behavioral change and the contribution of treatment-induced differences in misclassification rates.

Why Cross-Sectional Misclassification Rates Are Appropriate

My misclassification rates are measured at endline through DNA fingerprinting validation of farmers' self-reported varietal identification. While I also have baseline misclassification data, these baseline measures reflect identification accuracy for the variety portfolio farmers cultivated before the intervention. Because the treatment induced changes in which varieties farmers chose to grow, the sets of varieties being identified differ between baseline and endline, making direct temporal comparisons problematic.

For the purposes of this bounding exercise, cross-sectional comparison of endline misclassification rates between treatment and control groups is therefore the appropriate approach. The key question is not whether I can decompose misclassification changes over time, but whether treatment-induced differences in identification accuracy at endline could plausibly generate the observed consumption effects in the absence of behavioral change.

Because the randomized design ensures that treatment and control groups had no systematic differences in identification ability at baseline, any differences in misclassification rates observed at endline can be causally attributed to the intervention. These endline differences represent an upper bound on how much the treatment could have improved reporting accuracy, which is exactly what is required for assessing whether misclassification bias could explain away my results.

Finally, note that the treatment effect operates on the *intensive* margin of consumption, that is, on quantities or frequency of consuming orange or yellow varieties. On the extensive margin (whether households grow at least one such variety), the treatment shows no significant effect. This pattern suggests that misclassification

concerns primarily relate to how accurately households report consumption of varieties they already possess, not to false claims of consuming varieties they do not grow.

Misclassification Model

The relationship between reported consumption (C) and true consumption (C^*) follows the standard misclassification framework:

$$C = (1 - \delta) \cdot C^* + \gamma \cdot (1 - C^*),$$

where:

- δ is the false negative rate: Pr(Report No Trait | Actually Has Trait);
- γ is the false positive rate: Pr(Report Trait | Actually No Trait).

Decomposition of Treatment Effects

The observed treatment effect on reported consumption, $\hat{\beta}$, reflects both true changes in consumption behavior and potential differences in misclassification across treatment and control groups. In general, an exact decomposition includes interaction terms between behavioral change and reporting error. For the purposes of my bounding analysis, I focus on the null hypothesis of no true behavioral change, $C^*_T = C^*_C$. Under this restriction, the observed treatment effect simplifies exactly to:

$$\hat{\beta} = (\delta_C - \delta_T) \cdot C^* + (\gamma_T - \gamma_C) \cdot (1 - C^*)$$

where C^* denotes the common true consumption rate under the null. This expression isolates the portion of the estimated treatment effect that could attributed solely to differences in misclassification rates induced by the treatment. The bounding exercise then evaluates whether observed differences in false negative and false positive rates are large enough to account for the observed treatment effect in the absence of true behavioral change.

B. Bounding Exercise

To determine how large changes in misclassification rates must be to fully explain the observed treatment effect without behavioral change, I set $C^*_T - C^*_C = 0$:

$$\hat{\beta} = (\delta_C - \delta_T) \cdot C^*_C + (\gamma_T - \gamma_C) \cdot (1 - C^*_C).$$

This condition characterizes the combinations of false negative and false positive rate differences necessary to rationalize the observed effect through reporting behavior alone. Because the true consumption rate in the control group is unknown, I explore how required misclassification differences vary across all possible values of C_C^*.

Scenario Analysis

I examine three benchmark scenarios based on different assumptions about true consumption prevalence in the control group.

Scenario 1: Universal Baseline Consumption ($C_C^* = 1$) If all control households truly consumed orange or yellow varieties, the bounding condition simplifies to:

$$\hat{\beta} = \delta_C - \delta_T.$$

Here, only reductions in false negative rates could explain the observed effect. This is the most favorable scenario for the misclassification explanation, as it maximizes the role misclassification could play. If even under this extreme assumption the required change in false negatives is implausibly large, then misclassification cannot explain the results.

Scenario 2: Zero Baseline Consumption ($C_C^* = 0$) At the opposite extreme, if no control households truly consumed such varieties:

$$\hat{\beta} = \gamma_T - \gamma_C.$$

In this case, the entire observed effect would need to arise from increases in false positive rates. This assumption is maximally unfavorable to the misclassification explanation, as it would require the treatment to induce systematic over-reporting of consumption.

Scenario 3: Partial Baseline Consumption ($C_C^* \in (0, 1)$) The most realistic case allows for intermediate consumption prevalence:

$$\hat{\beta} = (\delta_C - \delta_T) \cdot C_C^* + (\gamma_T - \gamma_C) \cdot (1 - C_C^*).$$

For each value of C_C^*, I can compute the combinations of false negative and false positive rate differences that would be required to account for the treatment effect. This sensitivity analysis explores the full range of plausible misclassification adjustments.

C. Empirical Implementation and Interpretation

My DNA fingerprinting validation data provide a unique opportunity to assess the plausibility of the misclassification differences required to explain the observed treatment effects. For each scenario, I compare the required misclassification rate differences to those actually observed between treatment and control groups at endline.

Because households consume primarily from their own sweet potato production (93% from own harvest, 6% from market purchases), misclassification of production varieties provides a highly relevant proxy for consumption misreporting.

If the differences in misclassification rates required to explain away the treatment effect are much larger than those observed in the data, then misclassification cannot plausibly account for the results.

An additional supporting fact is that the treatment does not significantly affect the extensive margin of adoption (whether households grow at least one orange or yellow variety). This limits the scope for false positive reporting. Misclassification could matter only on the intensive margin, how accurately consumption of existing varieties is reported.

This empirical strategy does not rely on strong assumptions about the true consumption prevalence. Instead, it examines the full set of conditions under which the observed effects could be entirely driven by misclassification. Showing that these conditions are inconsistent with the empirical misclassification rates provides strong evidence that the intervention induced genuine behavioral change.

D. Results

Table A18 presents the results. Even under the most conservative scenario, assuming that all control households already consumed yellow-fleshed varieties in truth ($C_C^* = 1.00$), the required difference in false negative misclassification rates to fully explain the treatment effect would be 14.6 pp. However, the DNA fingerprinting validation data show that the observed difference in false negative rates between treatment and control at endline is only 5.3 pp, accounting for at most 36% of the observed treatment effect.

Similarly, explaining the effect purely through differences in false positive misclassification would require implausibly large gaps between treatment and control groups. Under intermediate assumptions about true baseline consumption ($C_C^* = 0.50$), the required difference in false positive rates would be 29.2 pp, whereas the observed difference is negligible at only 0.07 pp, less than 0.3% of what would be needed. Even in the most extreme scenario where no control households truly consumed yellow-fleshed varieties ($C_C^* = 0.00$), the required difference of 14.6 pp is more than 200 times larger

Table A18: Bounding Analysis for Bias from Differential Misclassification.

Assumed C_C^*	Required $\Delta\delta$	Observed $\Delta\delta$	Required $\Delta\gamma$	Observed $\Delta\gamma$
0.00		0.0533	0.1460	0.0007
0.25	0.5840	0.0533	0.1947	0.0007
0.50	0.2920	0.0533	0.2920	0.0007
0.75	0.1947	0.0533	0.5840	0.0007
1.00	0.1460	0.0533		0.0007

Notes: This table reports a bounding exercise assessing the extent to which differential misclassification between treatment and control groups could generate an observed estimated effect $\beta = 0.146$ through measurement error alone. Column (1) reports assumed values of the true outcome rate in the control group, denoted $C_C^* \in [0, 1]$. Columns (2) and (4) report the implied differences in misclassification rates required to produce β under these assumptions: the required difference in false negative rates is $\Delta\delta = \beta/C_C^*$ and the required difference in false positive rates is $\Delta\gamma = \beta/(1 - C_C^*)$. Columns (3) and (5) report the corresponding observed differences in misclassification rates estimated from the DNA validation data. $\Delta\delta$ is defined as the false negative rate in the control group minus the false negative rate in the treatment group, and $\Delta\gamma$ is defined as the false positive rate in the treatment group minus the false positive rate in the control group. Empty cells indicate undefined values when denominators equal zero.

than what I observe.

Across all plausible scenarios for true baseline consumption rates, the observed misclassification differences fall far short of the levels required to fully explain the 14.6 pp treatment effect through measurement error alone. At most, treatment-induced differences in varietal identification accuracy can account for approximately one-third of the observed effect under the most favorable assumptions.

F Appendix: Cost and Scaling Considerations

The intervention I study is more demanding than conventional extension. This section asks what it costs, what drives that cost, and which features would have to change for it to operate at scale.

Genotyping cost $10.45 per sample, with an additional $1.00 for DNA extraction and $1.00 for in-country sampling supplies; shipping the batch added $0.14 per sample. Allowing a further $2.00 to $3.00 per sample for incremental enumeration time, enumerator training, and supervision of the sample chain gives $12.60 to $15.60 per sample.[27] I collected 3.32 samples per tested household at baseline, so producing an individual card cost $42 to $52 per tested household. Ten of the twenty sampled households per village were tested, so the cost of the diagnostic averaged over sampled households is $21 to $26 per farmer.

This figure is the cost of the diagnostic alone. T1 and T2 delivered the same information session, and the identification and area results in Tables 4 and 6 are specific to T2: the session alone left both varietal identification and area allocation unchanged. T1 did raise consumption of yellow-fleshed varieties and shorten diarrhea episodes, so the diagnostic does not account for the full difference between T2 and the control group. What it purchases is the reallocation of area toward trait-advantageous varieties, and the larger and more consistent effects on consumption and child health reported in Tables 7 and 9. The session itself would be required under any information-based program and is not specific to this intervention; I do not price it here.

The first thing to note is a margin that cannot be cut. The intervention works by locating trait-advantageous varieties within a household's existing portfolio, and which varieties these are is unknown in advance, which is the premise of this paper. A protocol sampling only the variety occupying the largest share of a household's sweet potato area would frequently return a variety that is neither yellow-fleshed nor virus-resistant, leaving the farmer correctly informed but with no target to reallocate toward. Sampling the portfolio is what allows the card to rank the varieties a farmer already holds. The 3.32 samples per tested household therefore reflect the number of distinct varieties cultivated rather than redundancy in the protocol, and the cost per tested household is bounded below by varietal diversity in the sample rather than by the design.

Two features do govern how the cost would change with scale. The first is the share of households tested. The community card carried the same trait information to households that were not themselves tested, at negligible marginal cost, and the effects

[27] Costs are in 2022 U.S. dollars and reflect the sample order placed for this study. Card production costs were negligible. I exclude endline genotyping and the genotyping of control and T1 households, both of which served measurement rather than delivery.

I estimate for these community-information households are comparable in magnitude to those for individual-information households (Tables 10 and 11). Since the diagnostic is the only cost that scales with the number of households tested, a program testing a smaller fraction would deliver the same information at lower cost per farmer. How far this can be pushed is an open question: my design tested half of the sampled households, and I cannot speak to the performance of a substantially sparser design.

The second, and more consequential, is the coverage of the reference library. Of 3,316 samples collected, 1,641 matched a reference variety, and 1,286 matched a reference for which trait data were available. The cost per sample yielding usable information is therefore $32 to $40, roughly two and a half times the cost per sample collected. The shortfall has two components with different remedies. Approximately half of samples matched no reference, which would require genotyping additional landraces. A further 355 samples matched a reference for which trait data were not available; several of these references are among the most frequently encountered varieties in my sample. Characterizing them is a one-time expense, after which every subsequent farmer in the region can be advised without additional genotyping. Expanding library coverage therefore dominates expanding farmer coverage as a means of reducing the cost per unit of delivered information.

I exclude the cost of the existing library from these calculations. The trait characterization it embodies draws on breeding trial data accumulated over decades by national and international programs, and treating it as a recurring cost of farmer-level advice would misstate what a scaled program would need to spend. This is a choice rather than an accounting identity: a program operating in a crop or country without a comparable library would face the cost of building one. The relevant point is that where such investments already exist, they can be made usable for advice at the farmer level, which is not how they have generally been used.

One comparison places these figures in context. The HarvestPlus Reaching End Users project in Uganda cost $66 per targeted household, counting households reached directly and indirectly, or roughly $90 in 2022 dollars.[28] That project distributed and multiplied planting material, because orange-fleshed sweet potato was not locally available. My results indicate that trait-advantageous varieties are already present in these communities and are misidentified rather than absent, so that correcting beliefs is sufficient to reallocate area toward them. The comparison is between instruments rather than between programs of comparable scope: my figure omits the session, and theirs includes everything.

Finally, the figures above understate what the intervention delivers. The reductions

[28] de Brauw et al. (2018), Table 8, Panel A. The project ran through 2009; the base year of the cost accounts is not stated and I assume 2009.

in child diarrhea reported in Table 9 follow from the same expenditure that the adoption and production results already justify. These gains appear among precisely the children for whom dietary changes are documented, and not among older children in whose diets I observe no corresponding shift, which is what a food-based pathway would predict.