Farm Event Timeline Analysis (FETA): High-Frequency Farm Accounting for Peer Benchmarking

Evidence from a Pilot Deployment Among Smallholder Maize Farmers in Nigeria

Adam Abramson[1,2] and Toyin Samuel Olowogbon [1,2, 3]
[1]Department of Economics and Management, Faculty of Social Sciences and Humanities, Tel-Hai University of Kiryat Shmona and the Galilee, Upper Galilee, 1220800, Israel
[2]Migal – Galilee Research Institute, 1101600, Kiryat Shmona, Israel;
[3] Institute for Food Security and Commercial Agriculture, Federal University Oye-Ekiti,

Prepared for the NBER Conference on Economic Measurement in the Agricultural Economy, Washington, DC, November 6, 2026.

Acknowledgements: *This research was conducted as part of the Farm Event Timeline Analysis (FETA) project at MIGAL — Galilee Research Institute, in collaboration with government agricultural extension agents in Nigeria. The study protocol was approved by the Tel-Hai University Institutional Review Board (SSH-08-07-26). We are grateful to the participating field agents and farmers for their time.*

Research was supported by internal funding at the Migal Galilee Research Institute and in-kind support from the ARC AI program at Tel Hai University of Kiryat Shmona and the Galilee. The authors declare that they have no material financial relationships or other conflicts of interest related to the research described in this paper.

Abstract

Farm income and profitability are difficult to measure reliably. Many agricultural surveys rely on farmers to reconstruct production activities retrospectively, often weeks or months after they occur. This design compresses a season of input purchases, field operations, labor activities, harvests, and sales into a small number of aggregate responses, making it difficult to distinguish an accurate total from one affected by recall or reporting error. This paper introduces Farm Event Timeline Analysis (FETA), a lightweight measurement instrument that instead records discrete farm activities as dated events throughout the production season. A trained field agent conducts a short weekly phone interview with each farmer, recording individual input purchases, field operations, and labor tasks and anchoring activities to days after planting. Monetary and physical quantities are standardized per hectare, and leave-one-out peer means are calculated for each cost category, input dose, and price received, producing a farm-level record that can be benchmarked against comparable farms.

We report results from a first pilot deployment among 64 smallholder maize farmers across three Nigerian states—Benue, Ekiti, and Lagos—tracked weekly through one growing season by six field agents. The pilot provides two findings relevant to agricultural economic measurement. First, the event-level records reveal substantial heterogeneity in farm practices and outcomes: pooled across the sample, mean yield was 3.9 t/ha (SD 1.9), mean total production cost was ₦443,481/ha (US$327/ha; SD ₦449,639/ha), and mean net revenue was ₦1,316,122/ha (US$969/ha; SD ₦1,174,042/ha). Relative to the pooled leave-one-out peer mean, deviations in fertilizer, crop-protection, and other input expenditures, as well as yield, sales, and net revenue, are typically 65–100 percent. Restricting peer comparisons to farms within the same state reduces the dispersion in yield, revenue, and total cost by approximately one-half to two-thirds, while dispersion in fertilizer expenditure, nitrogen applied, and crop-protection inputs remains substantial. This pattern indicates that input-use practices vary considerably even among farms operating within the same region.

Second, the pilot illustrates the importance of preserving the temporal and spatial structure of agricultural data when interpreting input–outcome relationships. Mean yield ranges from 1.1 t/ha (SD 0.5) in Benue, to 4.3 t/ha (SD 0.4) in Ekiti, to 5.7 t/ha (SD 0.8) in Lagos, a gradient mirrored in net revenue per hectare, which was close to zero on average in Benue (₦−4,448/ha) with wide variance across individual farms (SD ₦104,069/ha), against ₦1,501,910/ha (SD ₦638,200/ha) in Ekiti and ₦2,347,720/ha (SD ₦1,042,515/ha) in Lagos. A random-forest permutation-importance analysis (mean cross-validated R^2 = 0.90, SD 0.07) attributes approximately 82 percent of explanatory power to region and planting date combined. Consistent with this regional structure, the pooled correlation between total cost and yield is positive (r = 0.53) but reverses sign within two of the three states (r = 0.66 in Benue; r = 0.13 in Ekiti; r = −0.38 in Lagos), illustrating a Simpson's-paradox reversal. The result cautions against interpreting pooled input-level correlations as evidence of causal input effects when substantial regional heterogeneity is present.

We conclude by considering how the FETA design could be extended to farm income, labor input, and field-level productivity measurement in other settings, including the United States. A companion endline survey currently underway will directly compare FETA-based measures with conventional retrospective recall. We also consider the principal constraint to scaling the current design—the weekly field-agent phone call—and propose replacing live interviews with farmer-recorded voice messages that can be transcribed and structured using large language models (LLMs) via chat platforms or automated calls. This offers a potential pathway for scaling event-based agricultural measurement while substantially reducing enumerator costs.

Keywords: *farm income measurement, peer benchmarking, high-frequency data, recall validation, smallholder agriculture, field-level productivity data*

1. Introduction

Farm income and profitability are central measures of agricultural performance, yet they are difficult to measure reliably. Unlike many economic outcomes that can be observed directly, farm income and profitability are constructed from information on production, input use, labor, prices, sales, and other farm activities. The quality of these measures therefore depends not only on the statistical methods used to aggregate them, but also on how the underlying information is collected. This issue is recognized across agricultural statistical systems. In the United States, for example, the Agricultural Resource Management Survey (ARMS) is the U.S. Department of Agriculture's primary source of information on farm production practices, resource use, and economic well-being. In low- and middle-income countries, household and agricultural surveys similarly provide much of the information used to measure smallholder production and productivity. Across these settings, improving the structure and quality of agricultural data remains an important methodological challenge (Carletto et al., 2021).

A central difficulty is that agricultural production is a sequence of events, while conventional survey instruments often measure it retrospectively as a set of seasonal aggregates. A farmer may purchase fertilizer on several occasions, undertake different field operations at different stages of crop growth, hire labor for particular tasks, harvest on multiple dates, and sell output through several transactions. A retrospective interview may nevertheless ask the farmer to report total fertilizer expenditure, total labor use, total harvest, or total revenue for the season. Once these activities have been compressed into aggregate responses, the researcher has limited ability to determine how the reported total was generated or to distinguish a genuine extreme observation from an omitted activity, recall error, or reporting inconsistency.

Evidence from agricultural survey experiments demonstrates that recall can generate systematic measurement error. Using Living Standards Measurement Study–Integrated Surveys on Agriculture data from Malawi and Tanzania, Wollburg et al. (2021) show that reported harvest, labor, fertilizer use, and the number of plots vary systematically with the length of the recall period. The effects are economically meaningful: the estimated recall effect is typically between 2 and 5 percent of the value of the variable for each additional month of recall. The problem is particularly evident for agricultural labor, where activities vary substantially across plots, workers, and stages of production. In a randomized survey

experiment in Tanzania, Arthi et al. (2018) found that respondents reporting labor retrospectively at the end of the season reported up to four times as many hours per person-plot as respondents reporting labor weekly. Evidence from Ghana similarly shows that end-of-season recall produced approximately 10 percent higher reported farm labor per person-plot, largely because recall respondents omitted marginal plots and workers (Gaddis et al., 2021). These findings demonstrate that recall error can affect both the quantity of reported activity and the composition of the underlying observations.

Measurement problems are not limited to labor or other frequently occurring activities. Even relatively stable quantities such as cultivated land area can be measured with systematic error. Comparing farmer-reported and GPS-based measures across African countries, Carletto et al. (2015) show that self-reported land area differs systematically from objective measurements and that these differences can affect estimates of farm size and agricultural productivity. More broadly, Abay (2020) demonstrates that measurement error in self-reported agricultural inputs and production can have important consequences for estimates of marginal returns to modern agricultural technologies. Abay et al. (2021) further show that the mechanism generating measurement error matters for economic inference: survey discrepancies may reflect both misreporting and farmers' perceptions of the underlying quantities on which they make decisions. Thus, improving measurement is not simply a matter of reducing random noise; the structure of the data-generation process can affect the economic relationships estimated from the data.

One response has been to collect information closer to the time when agricultural activities occur. Production diaries, for example, can reduce the period over which respondents must recall individual activities. Deininger et al. (2012) show that a production diary administered alongside a conventional household survey in Uganda generated systematically different estimates of output value from those obtained through retrospective reporting. Other innovations have used electronic data collection and computerized consistency checks to identify unusual or internally inconsistent observations. Fafchamps et al. (2012), for example, demonstrate that computerized checks can improve the precision of some reported measures of enterprise sales and profits. The broader agricultural measurement literature consequently emphasizes the importance of data structure—including questionnaire design, timing, measurement procedures, and technological implementation—for the reliability and usefulness of agricultural statistics (Carletto et al., 2021).

These approaches, however, leave a broader question unaddressed: shortening the recall period or validating a particular variable does not by itself preserve the sequence of dated events from which a cumulative farm statistic, such as total cost or total revenue, is actually constructed.

This paper introduces **Farm Event Timeline Analysis (FETA),** also referred to as the MicroLog instrument, as an event-based approach to agricultural economic measurement. FETA replaces a single end-of-season reconstruction with a weekly log of discrete, dated farm events: field agents record individual input purchases, field operations, and labor activities throughout the production cycle, with each activity anchored to days after planting.

The key methodological feature of FETA is consequently not simply higher-frequency data collection. It is the retention of the event history underlying the aggregate statistic. Individual expenditure observations

can be summed to reconstruct total input costs, and labor observations can be associated with particular field operations, creating a degree of traceability that is difficult to obtain from a conventional end-of-season questionnaire. It also permits the construction of leave-one-out peer benchmarks for individual farmers, allowing the researcher to examine how particular input-use practices and outcomes differ from those of comparable farmers.

Retaining each activity's date has a further, agronomic implication beyond measurement accuracy. Because every observation is anchored to days after planting rather than folded into a season total, the record preserves not only how much of an input a farmer applied but when, relative to crop development, and in what sequence relative to other operations. A seasonal aggregate cannot support this kind of analysis by construction: two farmers who apply identical total fertilizer quantities but at different growth stages, or who plant at different points in the rains, are indistinguishable in a single end-of-season total, yet timing of this kind is often as consequential for yield as the quantity applied. In the pilot reported here, for example, planting date turns out to be among the strongest single correlates of yield (Section 4.3), a relationship that only a timing-resolved record could have surfaced.

Because observations are also logged as they occur rather than reconstructed afterward, the same structure could in principle support near-real-time feedback to farmers during the season itself, once appropriate communication channels are in place: flagging an emerging problem, such as a delayed input application, while there is still time to respond, rather than only in retrospect. The present pilot does not implement such a feedback loop; Section 5.2 returns to this possibility directly.

We evaluate FETA using its first pilot deployment among 64 smallholder maize farmers across three Nigerian states. Six trained field agents followed participating farmers weekly over one growing season. The pilot was designed primarily to assess the feasibility of implementing the instrument under field conditions and to explore the information generated by its event-level structure, rather than to provide definitive estimates of Nigerian maize productivity. The resulting data allow us to examine the extent of farm-level heterogeneity in input use and outcomes and to assess how these patterns change when observations are benchmarked against comparable peers.

The paper makes two contributions. First, it demonstrates the feasibility of collecting a season-long, event-level record of smallholder agricultural activities using a relatively lightweight field-agent-administered instrument. Second, it demonstrates how these records can be used to characterize farm-level heterogeneity through standardized peer comparisons. In the pilot, fertilizer expenditure, crop-protection spending, and applied nitrogen show substantial dispersion across farms relative to their peers, while regional differences account for an important share of the variation in outcomes. We therefore use the pilot data to illustrate an additional analytical implication of high-resolution agricultural records: relationships between farm inputs and outcomes can be misleading when substantial spatial and temporal heterogeneity is not accounted for. The exploratory analysis highlights this issue through regional comparisons and a Simpson's-paradox reversal in the relationship between costs and yields.

Our contribution is therefore best understood as one of measurement architecture rather than measurement validation. We do not claim that the pilot establishes that MicroLogs eliminate measurement error or produce more accurate estimates than conventional surveys. Such a claim would require an external benchmark or experimental comparison with an alternative measurement protocol. Instead, we show that agricultural economic records can be structured so that the aggregate measures used in analysis remain connected to the individual events from which they are constructed, with two direct advantages. First, this makes it possible to inspect, benchmark, and reconstruct the components of farm-level economic outcomes rather than relying exclusively on a farmer's retrospective reconstruction of an entire production season. Second, because each event is dated rather than folded into a seasonal total, the record preserves the timing and sequencing of farm activities relative to crop development, information a seasonal aggregate discards by construction and that is directly relevant to agronomic questions about how the quantity and timing of an input, not simply its total amount, relates to yield.

This distinction is important for agricultural economic measurement more broadly, where recall length, reporting protocols, and survey design are already known to materially affect the resulting statistics (Carletto et al. 2021). FETA asks whether the underlying event structure of agricultural production can itself become part of the measurement record; the pilot reported here is a first step in evaluating that possibility at a scale that could support both conventional aggregate analysis and event-level investigation of how those aggregates were generated.

The remainder of the paper is organized as follows. Section 2 reviews the literature on recall bias and measurement error in agricultural surveys, high-frequency data collection, and alternative approaches to measuring agricultural economic activity. Section 3 describes the FETA instrument, the pilot's design and study setting, the analytical sample and data-processing procedures, and the study's ethical review. Section 4 presents the results, including farm-level deviations from peer benchmarks (Section 4.1), descriptive outcomes (Section 4.2), and an exploratory analysis of yield correlates (Section 4.3). Section 5 discusses the role of regional heterogeneity, including a Simpson's-paradox reversal in the cost–yield relationship (Section 5.1), and implications for agricultural economic measurement more broadly (Section 5.2). Section 6 concludes with directions for a larger confirmatory deployment.

2. Related Literature

This paper draws on three literatures: survey-methods research on recall bias and measurement error in agricultural data, research on high-frequency and phone-based data collection, and research on peer benchmarking, social comparison, and social learning as mechanisms that shape what farmers know and do. We review each in turn and then position the paper's contribution relative to them.

Recall bias and measurement error in agricultural surveys

The premise of this paper, that a single end-of-season recall interview is a poor instrument for measuring farm costs, labor, and output, rests on a well-established empirical literature. Using LSMS-ISA panel data from Malawi and Tanzania, Wollburg, Tiberti, and Zezza (2021) show that key input and output variables are systematically related to the length of the recall period, with reporting error increasing by

approximately 2 to 5 percent for each additional month of recall length, indicating a nonrandom bias rather than noise that would average out across a sample. Arthi et al. (2018) extend this finding using a randomized design: farm households in Tanzania were assigned to either a single post-season recall module or a weekly labor diary covering the same season, and the recall module overstated labor hours by a factor of approximately 3.4, with error at both the intensive margin (hours) and the extensive margin (which plots and household members were active) moving in economically consequential directions. Measurement error of this kind is not confined to flow variables such as labor or expenditure. Carletto, Gourlay, and Winters (2015) and Abay, Bevis, and Barrett (2021) show that even self-reported plot area, a stock variable a farmer might be expected to know precisely, carries systematic, non-classical error correlated with plot size itself, large enough to bias estimated productivity relationships, including the long-debated inverse relationship between farm size and yield, when used without instrumentation. Taken together, this literature establishes both that recall length drives measurement error and that the resulting error is structural rather than idiosyncratic. The weekly, event-anchored design of FETA is a direct response to the first finding; its use of a plot size measured once, rather than re-estimated from memory each week, for per-hectare standardization is a partial response to the second. This literature does not, however, provide an instrument that allows a single deployment to assess its own accuracy against an independent farmer estimate. Closing that gap is the purpose of the endline survey described in Section 6.

High-frequency and phone-based data collection

A separate literature examines whether frequent, phone-mediated contact can substitute for periodic in-person interviews. The clearest large-scale precedent is the World Bank's LSMS-supported COVID-19 High-Frequency Phone Survey program, fielded on a monthly basis across more than 100 countries beginning in 2020. Gourlay et al. (2021) summarize the resulting good practices and open questions, including tradeoffs that any phone-based instrument, including FETA, must design around: under-coverage of households with limited connectivity, elevated non-response and attrition, and a practical ceiling on interview length. On the extension-delivery side, Fabregas, Kremer, and Schilbach (2019) synthesize a set of randomized trials testing SMS-based agricultural advice across East Africa and find that low-cost text messaging alone had limited effects on farmer knowledge and input adoption relative to richer, in-person channels such as farmer field days, a useful caution that the details of channel design matter and that a low-cost digital channel is not automatically an effective one. That literature evaluates phone-based channels as a vehicle for persuasion and advice, whereas FETA uses a phone-based channel for measurement. Section 5.2 draws on the same underlying lesson, that message richness and respondent burden matter, in proposing farmer-recorded voice messages, rather than a live structured interview, as a subsequent step in this design space.

Peer benchmarking, social comparison, and social learning in agriculture

A third literature motivates why a peer-benchmarked dashboard, rather than a private record for the farmer or an internal analytic dataset for the researcher, is a natural output of an instrument of this kind. Conley and Udry (2010) show that Ghanaian pineapple farmers adjust their own input use toward that of their “information neighbors” after observing those neighbors' outcomes, establishing that farmers treat peer input-use information as decision-relevant rather than merely descriptive. BenYishay and Mobarak

(2019) show that the persuasiveness of agricultural information depends on the perceived similarity between the source and the recipient, with farmers more responsive to peers facing comparable conditions and constraints. This finding provides the empirical basis for the paper's argument, developed in Sections 4.1 and 5.1, that a peer-benchmarking dashboard should compare a farmer against a regionally comparable peer group rather than a pooled national sample. Kondylis and Jones (2018) provide a causal estimate of the value of feedback itself in an agricultural-services setting: eliciting and returning feedback increased farmer demand for the underlying service, with positive spillovers to nearby non-clients. Outside agriculture, Allcott (2011) offers the canonical demonstration of the mechanism that the FETA dashboard implements directly: comparing an individual's own resource use against a peer average reduces that use, in Allcott's case for residential electricity consumption, delivered at national scale because household smart meters made a rich, individualized peer comparison inexpensive to construct. Smallholder and mid-size U.S. farms have no equivalent automatic sensor network; the FETA activity log functions, in effect, as the infrastructure that would need to exist before an Allcott-style peer comparison could be constructed in this setting.

Positioned against these three literatures, the contribution of FETA is to operationalize the weekly-logging response that the recall-bias literature calls for, using the lower-cost phone-based channel that the high-frequency-survey literature has shown to be logistically viable, and to render leave-one-out peer benchmarking, the mechanism the social-comparison literature identifies as behavior-changing, a standard output of the data-collection process itself rather than a separate feedback intervention layered onto conventionally collected data, as in Kondylis and Jones (2018) or Allcott (2011). To our knowledge, this combination of event-level input logging and automatic, standardized peer benchmarking within a single low-cost instrument has not previously been evaluated in the agricultural-measurement literature, which is the gap this pilot begins to address.

3. Methodology

3.1. The FETA Instrument

Farm Event Timeline Analysis (FETA), also referred to as the MicroLog instrument, is a structured, spreadsheet-based data-collection system designed to record agricultural production activities as discrete events over the course of a growing season. The instrument is administered weekly by a trained field agent through a short phone interview with the farmer. Each field agent operates an individual workbook, and observations are accumulated throughout the production cycle rather than reconstructed retrospectively at the end of the season.

The instrument has three linked components: (i) an activity-level cost and input log, (ii) a weekly crop-condition survey, and (iii) standardized farm-level measures and peer benchmarks.

3.1.1. Activity-level cost and input log

The core component records each discrete farm activity as an individual dated observation. Activities include land preparation, planting, fertilizer application, herbicide application, weeding, pest management, and harvesting. For each activity, the instrument records the input or material used,

quantity, unit cost, and labor inputs. Labor is recorded separately for hired and family labor, with hours assigned to the relevant activity.

The event-level structure is cumulative: each new weekly interview adds observations to the existing farm record. Consequently, seasonal totals are constructed by aggregating the underlying activity records rather than by asking the farmer to report a single retrospective total. The same records can therefore be disaggregated by activity, input category, or timing and subsequently compared across farms. This structure also provides the basis for the peer-benchmarking analysis described below.

3.1.2. Weekly crop-condition survey

The activity log is supplemented by a short weekly condition survey administered during the same phone call. The survey records a crop-condition rating on a 1–5 scale, rainfall and soil-moisture conditions, drainage problems, and farmer-reported shocks. The shock field is open-ended and can capture events such as pest outbreaks, dry spells, and flooding.

Each activity and condition-survey observation is anchored to days after planting (DAP). The resulting data form a season-long timeline that links farm activities to the stage of crop development at which they occurred. This temporal structure allows the instrument to retain information about the sequencing of production activities that would be lost when the season is summarized only through end-of-season aggregates.

3.1.3. Standardization and peer benchmarking

To facilitate comparison across farms of different sizes, monetary values, nutrient quantities, and yields are standardized on a per-hectare basis using each farm's recorded plot area. Nutrient quantities are calculated from the fertilizer products recorded in the activity log using a fertilizer-product lookup table. The standardized nutrient measures include nitrogen (N), phosphorus pentoxide (P_2O_5), and potassium oxide (K_2O).

For each farmer (i), we calculate a leave-one-out peer mean for each cost category, nutrient measure, and price received per unit sold. The leave-one-out mean excludes the focal farmer from the benchmark:

For each farmer i, the leave-one-out peer mean for a given measure is:

$$\bar{X}_{-i} = 1/(N-1) \times \Sigma_{j=1, j\neq i}^{N} X_j \qquad (1)$$

where $\bar{X}_{-i}$ is the leave-one-out peer mean for farmer i; X_j is the corresponding value for farmer j; N is the number of farmers with a valid observation for the specific measure; and $j \neq i$ indicates that farmer i is excluded from the peer group.

In the FETA application, the calculation is repeated separately for each standardized measure, such as fertilizer expenditure per hectare, nitrogen applied per hectare, crop-protection expenditure per hectare, or price received per unit sold.

These peer means are used to construct farm-level benchmarks for input use, production costs, prices, and outcomes. The resulting information is presented in a one-page-per-topic dashboard that compares each farmer's standardized measures with the corresponding peer benchmark and displays the farm's activity timeline. The dashboard is designed so that information collected during the production season can, in principle, be returned to the farmer or an extension agent within days of collection rather than only after completion and processing of a season-end survey. (Figure 1)

Figure 1. Sample Peer-Benchmarked Economics Summary

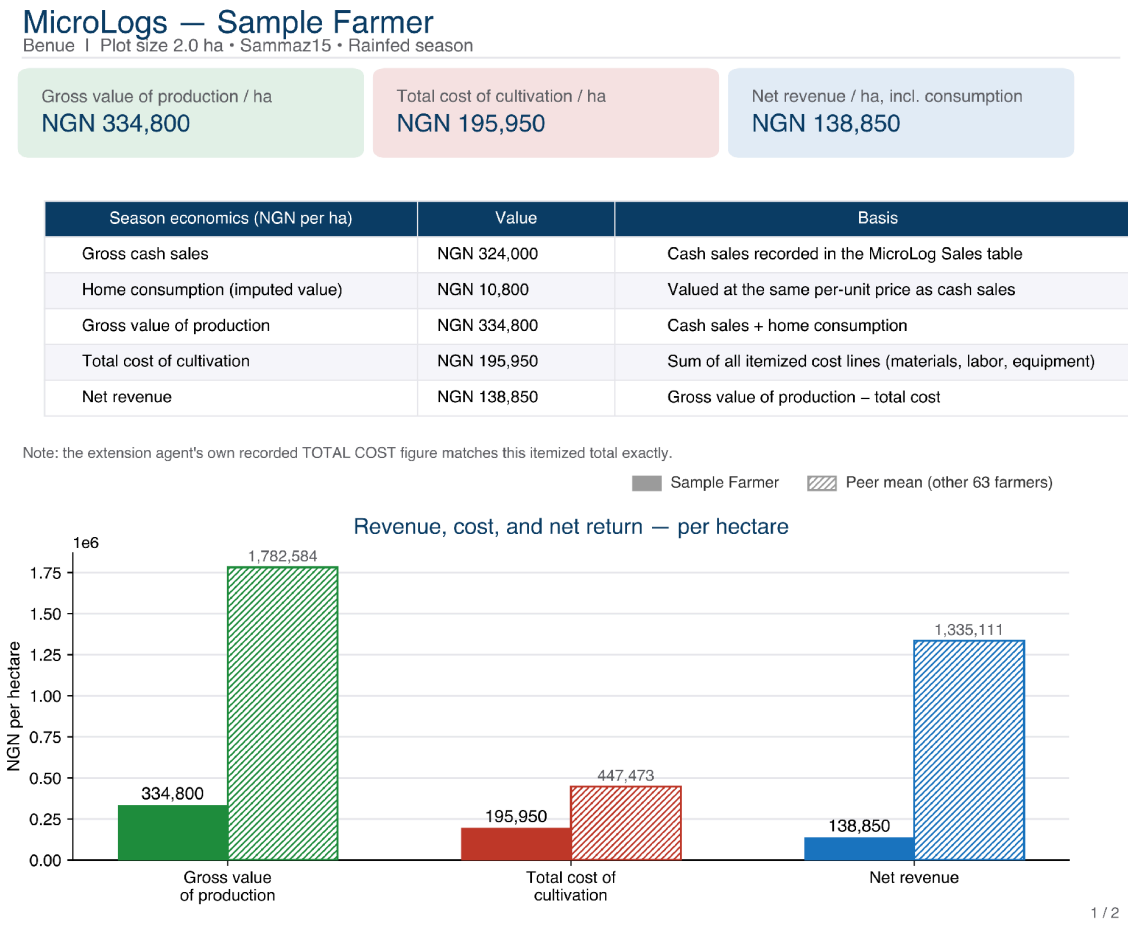


Figure 1. Example FETA output: one farmer's per-hectare economics summary, cost breakdown, and leave-one-out peer comparison (filled markers = this farm; hollow markers = peer mean). Illustrative example from the pilot sample.

Figure 2. Sample Days-After-Planting Activity Timeline

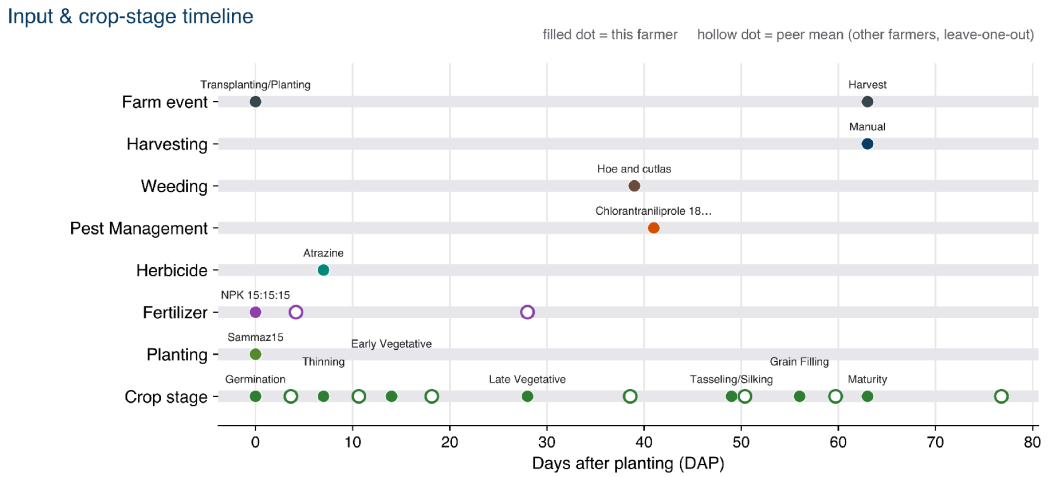


Figure 2. The same farm's days-after-planting activity timeline, with peer-mean timing markers overlaid.

3.2. Study Site and Pilot Design

Maize is one of Nigeria's principal staple and cash crops, and Nigeria has recently become Sub-Saharan Africa's largest maize producer by volume, a position built overwhelmingly on smallholder, rainfed cultivation rather than large-scale commercial farming (Chiaka et al. 2022). National average yields remain low relative to attainable levels, and productivity varies markedly across Nigeria's agro-ecological zones, from the extensive Guinea Savanna grain belt of the north-central states to the smaller-holding, more market-connected farming systems of the southwest (Chiaka et al. 2022; Sheahan and Barrett 2017). This pilot was deliberately located to span that variation.

The pilot was implemented during one maize growing season in 2026 across three Nigerian states: Benue in north-central Nigeria and Ekiti and Lagos in southwestern Nigeria (Figure 3). States were selected for logistical reasons related to the availability and location of the field-agent network rather than through a probability-based sampling procedure.

Figure 3. Study States Within Nigeria

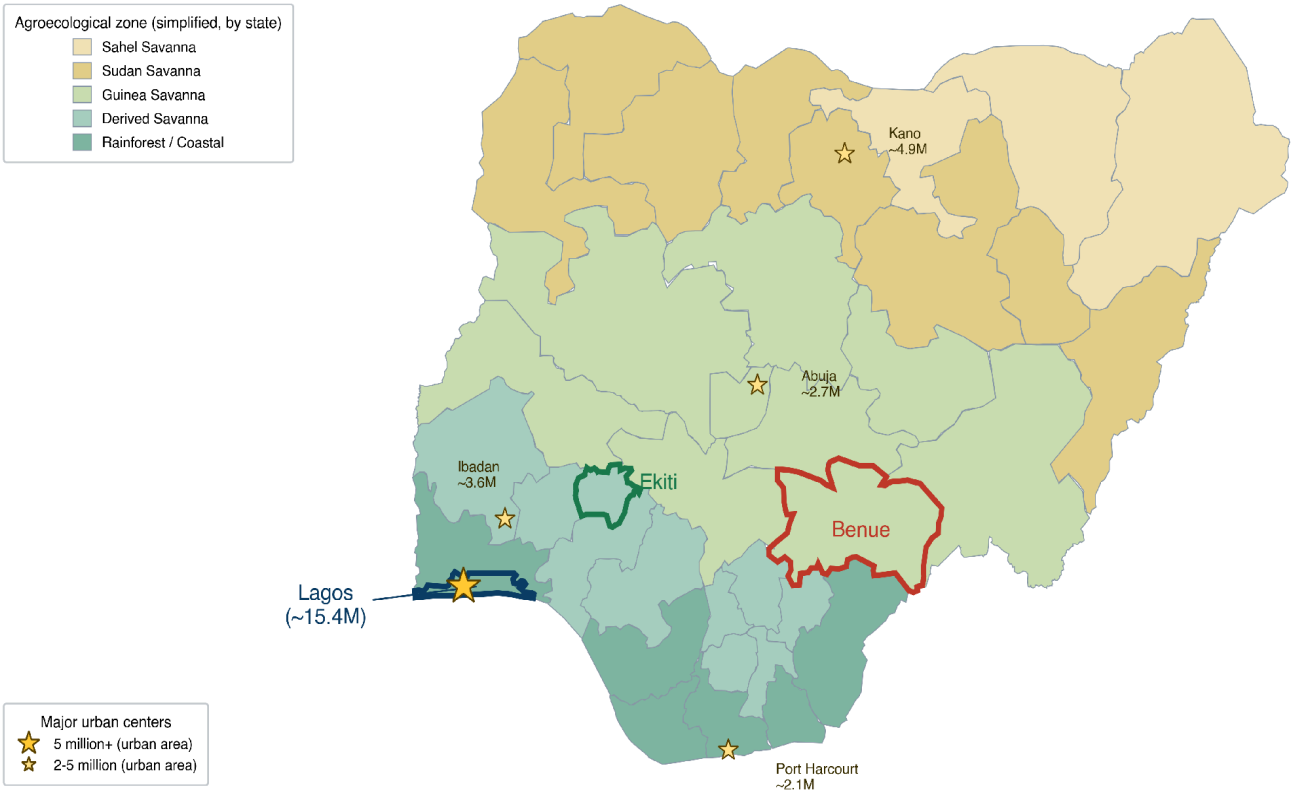


Figure 3. The three states covered by the FETA pilot: Benue (north-central Guinea Savanna) and Ekiti and Lagos (southwest).

Field agents in this pilot were drawn from Nigeria's state-level Agricultural Development Programme (ADP) extension network, the public system through which each state government delivers agricultural advisory services to farmers. The ADP structure was established with World Bank support beginning in the 1970s and 1980s and is funded jointly by federal, state, and donor sources; within each state ADP, extension agents are assigned to defined geographic cells and report through a supervisory hierarchy up to a state director of extension services (Auta and Dafwang 2010; Inegbedion et al. 2019). This existing, government-recognized point of contact between extension services and smallholder farmers is what the pilot's field-agent network builds on, rather than a purpose-recruited enumerator team unfamiliar to respondents.

Six trained field agents independently administered the MicroLog instrument to between 10 and 12 farmers each. Each field agent operated within a single state. The resulting sample comprised 64 farmers: 20 in Benue, 24 in Ekiti, and 20 in Lagos (Table 1).

Table 1. Field agents, their assigned state, and number of farmers surveyed

Field agent	State	Farmers (n)
Agent A	Benue	10
Agent B	Benue	10
Agent C	Ekiti	12
Agent D	Ekiti	12
Agent E	Lagos	10
Agent F	Lagos	10
Total	3 states, 6 agents	64

The pilot constitutes a convenience sample rather than a probability sample. The results should therefore not be interpreted as population estimates of Nigerian maize production, input use, or profitability. The purpose of the deployment was to assess the feasibility of implementing the FETA instrument under field conditions and to generate an initial standardized panel of event-level farm records for examining farm-level heterogeneity and the properties of the measurement system.

An endline survey of this same sample is being conducted as this paper is being prepared and is not yet complete, so its results are not reported here. Rather than relying solely on the season-long MicroLog log, the endline instrument returns to each farmer at the close of the season and elicits, before showing any calculated figure, the farmer's own unaided estimate of total cost, revenue, and profit for the reference plot; it then presents a line-by-line comparison against the corresponding totals computed from that same farmer's MicroLog activity log and asks directly which figure the farmer trusts more and why. A separate module reviews the farmer's own FETA activity timeline alongside an anonymized average timeline across participating farmers, asking whether the timeline correctly represents the season and whether the peer comparison itself is useful, and a closing module assesses the farmer's experience with the weekly MicroLog process, including perceived burden and willingness to continue this kind of record-keeping in a future season. The endline is designed to speak directly to a question the pilot-season data alone cannot answer: how closely a farmer's own recall and perception of season-level profitability track the event-level MicroLog record, and whether farmers themselves find the peer-benchmarked feedback useful. We plan to report these results in a follow-up analysis.

3.3. Analytical Sample and Data Processing

Of the 64 enrolled farmers, 59 have a usable yield observation. Five records contain missing or unparseable harvest quantities and are therefore excluded from analyses requiring yield. All 64 farmers have at least a partial cost log. Sixty-three farmers have usable measures of total cost and net revenue per hectare. One farmer's plot area is missing; this observation is retained where possible for non-area-standardized measures but is excluded from analyses requiring per-hectare standardization rather than having the missing area imputed.

The event-level structure of the instrument also facilitated the identification and correction of several data-entry and template artifacts during data processing. These were distinguished from farmer-reported

measurement error because they arose from the data-collection template or entry process and could be identified through the internal structure of the event records.

First, 20 of the 64 records contained implausible activity dates inherited from an unedited template default. These dates were excluded from the season timeline. Second, in 12 of the 64 records, a harvest quantity's monetary value had been entered in a cost column rather than the revenue column. The affected records were corrected consistently across the relevant field-agent workbooks. Third, two records contained sales totals that were inconsistent with the corresponding recorded quantity and price. These were corrected using farmer-specific overrides supported by the underlying records.

These corrections were possible because individual events contain independent information on date, unit, quantity, price, and the role of the observation within the farm account. The data-processing procedure therefore allowed implausible or internally inconsistent records to be identified through predefined plausibility and consistency checks. Importantly, none of these corrections required comparison with an independently elicited farmer-reported seasonal total. The corrections instead relied on the internal structure of the event-level record itself.

3.3.1 Data Quality and Internal Consistency

The event-level structure of the instrument permits several checks on data completeness and internal consistency beyond the season-level summary statistics reported above. Across the 64 farmers, field agents logged 1,493 discrete activity entries, an average of 23.3 per farmer (SD 6.9; range 11–43), spanning a mean of 6.2 of the instrument's activity categories per farmer (SD 1.5; range 3–10 of land preparation, planting, fertilizer, herbicide, weeding, pest management, and harvesting). Weekly crop-condition surveys were similarly complete: farmers were tracked for a mean of 12.0 of the 12 scheduled survey weeks (SD 0.3; range 10–12), and all 64 farmers have a recorded date for at least 7 of the instrument's 9 defined growth stages (mean 7.8). All 64 farmers have at least a partial cost log, and 63 have usable per-hectare measures of total cost and net revenue (one farmer's plot area was not recorded and is excluded from per-hectare standardization). During processing, the event-level structure also allowed systematic data-entry issues to be identified and corrected rather than left as unexplained noise: 20 of 64 records contained an implausible activity date traceable to an unedited template default, and 12 of 64 records had a harvest value entered in the wrong column (cost instead of revenue), both corrected consistently using the underlying event log rather than discarded. The near-complete weekly survey coverage and the ability to trace and correct the great majority of flagged anomalies back to identifiable, systematic causes — rather than to unrecoverable respondent error — is itself evidence that the event-level design supports meaningful data-quality auditing at a granularity a season-end recall survey could not provide.

3.4 Ethical Considerations

The study protocol was reviewed and approved by the Tel-Hai University Institutional Review Board (IRB) (Protocol No. SSH-08-07-26). Prior to participation, all respondents were provided with information about the study, including its purpose, the nature of their participation, the information to be

collected, and their rights as research participants. Informed consent was obtained from each participant before data collection commenced. Participation was voluntary, and respondents were informed that they could decline to answer any question or withdraw from the study at any time without penalty or loss of any benefits to which they were otherwise entitled. Data were collected and handled in accordance with the confidentiality and data-protection procedures specified in the approved study protocol.

4. RESULTS

4.1 Peer-Benchmarked Variation in Farm Practices and Outcomes

The same leave-one-out peer means that populate each farmer's dashboard (Figure 1) allow us to characterize, across the full pilot, how far an individual farm's practices and outcomes typically sit from a comparable peer average. This is a direct measure of the informational content a peer-benchmarking instrument such as FETA can provide to an individual farmer. For seven key categories, we compute each farm's percentage deviation from the leave-one-out mean of all other farms, first pooling across the full 64-farm sample and then restricting the peer group to farms in the same state. Table 2 and Figure 4 summarize the resulting distributions.

Table 2. Median absolute percentage deviation from the leave-one-out peer mean, by category and comparison scope (pooled vs. within region).

Category	Pooled median deviation	Within-region median deviation
Fertilizer cost / ha	100%	88%
Herbicide cost / ha	88%	71%
Pest management (sprays) cost / ha	100%	100%
N applied (kg/ha)	100%	100%
Gross sales value / ha	87%	25%
Total cost / ha	65%	33%
Net revenue / ha	87%	33%

Figure 4. How Far Individual Farms sit from their Peer Average, by activity

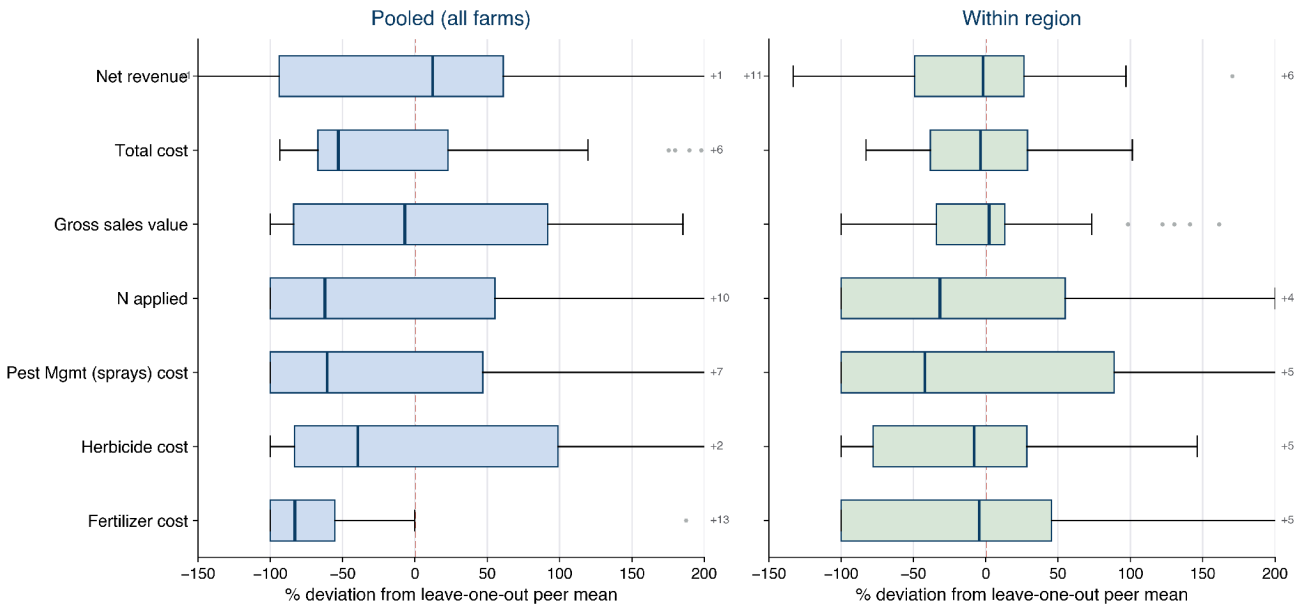


Figure 4. Distribution of individual farms' percentage deviation from their leave-one-out peer mean, by category, pooled (left) versus within region (right). Outcome categories (net revenue, total cost, gross sales value) compress substantially moving from the pooled to the within-region comparison; input and practice categories (fertilizer, herbicide, pest management, nitrogen) do not.

Two patterns are apparent in Table 2. First, pooled deviations are large across every category: a typical farm's total cost, net revenue, or fertilizer spending sits 65 to 100 percent away from the pooled peer mean. This finding is substantially a restatement of the regional pattern documented in Sections 4.2 through 5.1, since Benue, Ekiti, and Lagos differ so markedly on average that a farm compared against the full pooled sample is largely being compared against farms operating in a different agro-ecological and market context. Second, and more informative, restricting the peer group to the farmer's own state substantially reduces the outcome-side deviations: median deviation falls from 87 to 25 percent for gross sales value, from 87 to 33 percent for net revenue, and from 65 to 33 percent for total cost. It does not, however, meaningfully reduce the input-side deviations: fertilizer cost declines only from 100 to 88 percent, while nitrogen applied and pest-management spending are both unchanged at 100 percent. These input-side deviations remain nearly as large among farmers in the same state as across the full sample.

This pattern indicates that, once region is accounted for, farms in the same state converge substantially on similar outcomes but not on similar practices: two farmers operating under broadly similar conditions in the same state routinely differ considerably in how much they spend on fertilizer and crop protection. We interpret this as direct evidence of the kind of farm-level heterogeneity that a peer-benchmarked dashboard is intended to detect. It also carries a design implication for the instrument itself. Because a pooled peer comparison largely reflects the between-region gap documented in Section 5.1 rather than

the more actionable within-region variation in practice that a farmer or extension agent could act on, a production version of the dashboard should compute peer means within a comparable stratum, at minimum region, rather than pooling across a sample that spans multiple regions.

4.2. Descriptive, Peer-Benchmarked Results

Because every farm's cost, nutrient, and yield figures are standardized per hectare, they are directly comparable across the pilot despite plot sizes ranging from 0.5 to 4.0 hectares. Pooled across the three states, mean maize yield was 3.9 t/ha (SD 1.9; range 0.6–7.5 t/ha). Mean total production cost was ₦443,481/ha (US$327/ha), with a standard deviation of ₦449,639/ha (US$331/ha). Mean net revenue was ₦1,316,122/ha (US$969/ha) (SD ₦1,174,042/ha [US$865/ha]), with substantial variation across farms, ranging from a net loss of ₦831,933/ha (−US$613/ha) to a net return of ₦4,216,300/ha (US$3,106/ha).[1] The dispersion is large enough that pooled means obscure more than they reveal; Table 3 and Figure 5 disaggregate outcomes by state.

Table 3. Maize yield by state, among farmers with a valid yield observation.

State	Farmers with valid yield	Mean yield (t/ha)	SD	Range
Benue	16	1.07	0.49	0.60 – 1.88
Ekiti	24	4.31	0.35	3.80 – 5.33
Lagos	19	5.72	0.79	3.30 – 7.50

Figure 5 shows that yield, total cost, and net revenue per hectare all rise sharply and consistently along the same ordering across all three states: Benue lowest on every measure, Ekiti intermediate, and Lagos highest, with little overlap between farms in different states despite substantial farm-to-farm variation within each. Average net revenue per hectare in Benue was close to zero (₦−4,448/ha, US$−3/ha — effectively break-even on average, against a large standard deviation across individual farms), compared with ₦1,501,910/ha (US$1,106/ha) in Ekiti and ₦2,347,720/ha (US$1,730/ha) in Lagos. This ordering tracks the same agroecological gradient discussed in Section 5.1: Benue sits in the Southern Guinea Savanna, Ekiti in the Derived Savanna transition zone, and Lagos in the Rainforest/Coastal zone, a gradient that also corresponds to increasing annual rainfall from north to south — roughly 1,300 mm around Benue's Makurdi, 1,400 mm around Ekiti's Ado-Ekiti, and 1,700–1,800 mm in Lagos (NiMet, 2023) — with a longer and more reliable rainy season closer to the coast. Because rainfall, agroecological zone, and market access all move together across these three states, Figure 5 alone cannot separate their individual contributions to the regional pattern; Sections 4.3 and 5.1 return to this identification problem directly.

[1] US-dollar equivalents are calculated using the Nigerian Foreign Exchange Market (NFEM) rate of ₦1,357.61 per US$1 on August 17, 2026.

Figure 5. Yield, Cost, and Net Revenue by State

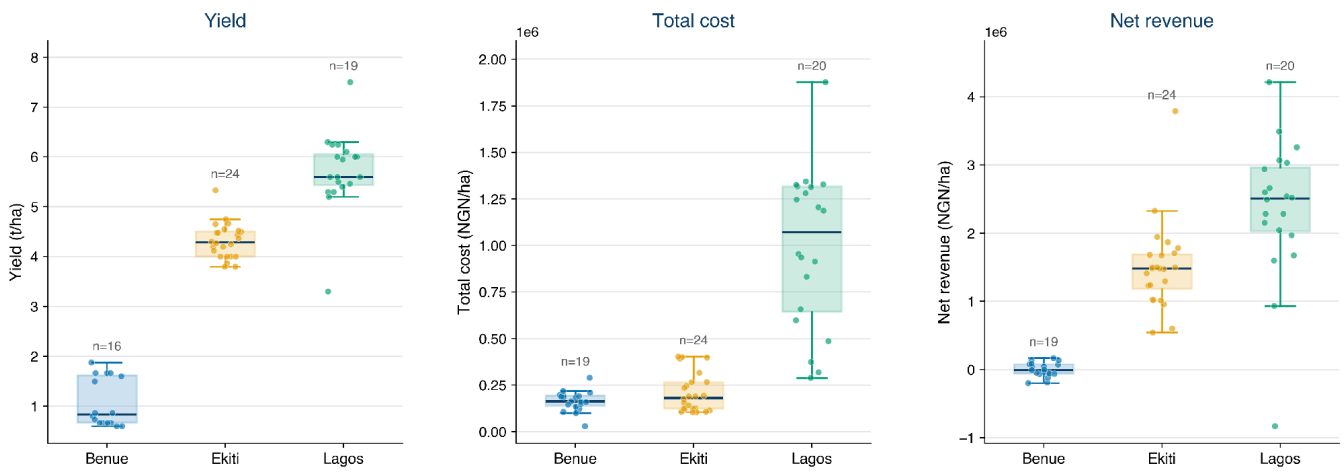


Figure 5. Yield, total cost, and net revenue per hectare by state, with individual farms shown as points (n as annotated). Benue's yields and costs are systematically lower than those of Ekiti and Lagos.

This standardized, peer-benchmarked structure constitutes a measurement product distinct from the yield analysis in Section 4.3: it is what allows an extension agent to convey to an individual farmer, in the farmer's own units, how their fertilizer cost per hectare or price received compares to the local peer average, a comparison that a one-time recall survey, processed in aggregate months later, cannot support.

4.3. Exploratory Correlates of Yield

We next used the standardized panel for an exploratory analysis of the correlates of yield, subject to the explicit caveat that this is a single-season, non-experimental, 59-farm sample. The results reported below are descriptive and exploratory rather than causal, and are reported because the pattern they reveal has a substantive methodological implication, developed in Section 5.1.

Table 4 reports univariate Pearson correlations between yield and each candidate predictor. Planting date (day of year) is the strongest single correlate ($r = -0.82$, $p < 10^{-15}$), followed by planting cost per hectare ($r = 0.78$), the number of weeks with heavy rain reported ($r = 0.66$), and herbicide cost per hectare ($r = 0.63$). Total cost per hectare ($r = 0.53$) and applied nutrients (nitrogen, phosphorus pentoxide, and potassium oxide, all $r \approx 0.45$–0.47) are more moderately correlated.

Table 4. Univariate Pearson correlations between yield and candidate predictors.

Predictor	r	p	n
Planting date (day of year)	−0.82	<0.001	59
Planting cost / ha	0.78	<0.001	58
# weeks with heavy rain	0.66	<0.001	59
Herbicide cost / ha	0.63	<0.001	58
Total cost / ha	0.53	<0.001	58
K_2O applied (kg/ha)	0.47	<0.001	59
P_2O_5 applied (kg/ha)	0.47	<0.001	59
N applied (kg/ha)	0.45	<0.001	59
Pest management cost / ha	0.44	<0.001	58
Fertilizer cost / ha	0.43	<0.001	58

We next estimate a cross-validated random forest (5-fold, 20 repeats), a nonparametric method that has performed well in comparable crop-yield prediction settings by capturing nonlinearities and interactions among agronomic predictors (Jeong et al. 2016). Using all candidate predictors together with region and variety fixed effects, the model achieves a mean out-of-sample R^2 of 0.899 (SD 0.074). Permutation importance attributes close to two-thirds of this explanatory power to a single region indicator (Benue, importance 0.58) and a further quarter to planting date (0.24); every agronomic input variable, including cost, nutrients, labor, and crop-condition ratings, individually contributes less than 0.02. Dropping region and variety entirely and re-estimating the model on agronomic and management variables alone yields an almost identical cross-validated R^2 of 0.885 (SD 0.080), driven overwhelmingly by planting date alone (Figure 6). In this pilot, therefore, region and planting date jointly account for nearly all of the cross-validated variance in yield; every input-quantity and cost variable examined here is, at best, a secondary correlate once those two factors are accounted for.

Maize variety is likewise correlated with region in this sample. We interpret this as an expected pattern rather than a data-quality concern, since improved maize varieties in Nigeria are bred and recommended for specific agro-ecological zones; a given region's farmers clustering on the varieties suited to that zone is therefore the expected outcome rather than an artifact of data collection.

Figure 6. Random-Forest Yield-Predictor Importance

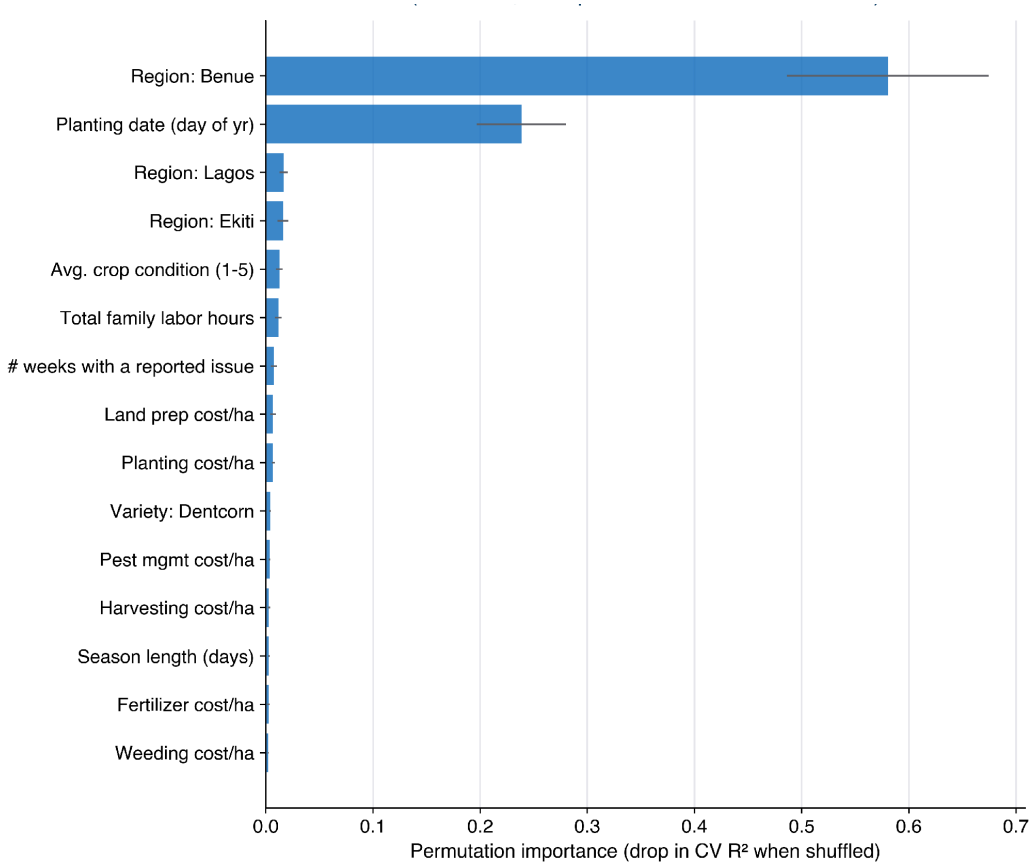


Figure 6. Random-forest permutation importance for yield, all predictors including region and variety fixed effects. Region (Benue) and planting date dominate; every input/cost variable contributes comparatively little.

An ordinary least squares regression with region fixed effects alone explains $R^2 = 0.914$ of the variance in yield; adding nitrogen applied, total cost per hectare, and the number of distinct fertilizer doses raises R^2 only to 0.924, and none of the three added agronomic terms is individually significant at the 5 percent level ($p = 0.086$–0.936). Section 5.1 discusses why this pattern warrants particular caution.

5. Discussion

5.1 Reading the Regional Pattern

The regional differences documented in Sections 4.2 and 4.3 are large, and they are consistent with what project field staff observed directly during the season: known agro-ecological and input-market differences across Benue, Ekiti, and Lagos are well understood by those who implemented the pilot, and

we have no basis for doubting the regional pattern itself. This is also consistent with the broader literature on regional heterogeneity in Nigerian smallholder agriculture. Benue lies in Nigeria's North Central geopolitical zone, part of the Guinea Savanna belt that functions as an extensive, high-volume grain-production zone, in contrast to the more urbanized, higher-market-access South West zone that includes Lagos and Ekiti (Chiaka et al. 2022). Farm mechanization and input intensity in Nigeria follow a similar north–south gradient: land-extensive, comparatively low-mechanization production is more typical in the north, while southern Nigeria has seen relatively greater mechanization and land-expansion pressure tied to input-intensive cultivation (Takeshima and Lawal 2018). Fertilizer profitability for maize in Nigeria is also substantially constrained by transportation costs and low marginal yield response, a constraint that bears more heavily on farmers further from input and output markets (Liverpool-Tasie et al. 2017). More generally, within-country heterogeneity in modern input use of this kind is a well-established feature of Sub-Saharan African agriculture, including Nigeria specifically (Sheahan and Barrett 2017). Taken together, these sources are consistent with the lower average input spending and yields we observe in Benue relative to Ekiti and Lagos (Table 3), rather than suggesting our regional pattern is an artifact specific to this pilot's field-agent assignment.

One methodological limitation is worth naming directly: because each of the six field agents in this pilot operated within a single state (Table 1), this dataset alone cannot formally separate a region effect from an enumerator effect, a concern with real empirical precedent — enumerator identity alone has been shown to account for a substantial share of variance in survey and assessment outcomes elsewhere (Rodriguez-Segura and Schueler 2023). We do not believe this changes the qualitative interpretation of the regional pattern reported here, but a crossed field-agent assignment in a larger follow-on deployment (Section 6) would let a future round confirm that formally.

A second, narrower caution concerns how the pooled correlations in Table 4 should be read. Because total cost and nitrogen applied both correlate with region, their pooled relationship with yield partly reflects the same regional gradient documented above rather than a within-state input–yield effect: state-specific correlations differ enough in sign and magnitude (Table 5; Figure 7) that a formal interaction test rejects a common slope across states for total cost ($p = 0.028$), though not for nitrogen ($p = 0.490$). None of the cost or nutrient correlations in Section 4.3 should therefore be read as evidence that additional spending or fertilizer raises yield, once region is taken into account.

Table 5. Pooled versus state-specific correlations between yield and key inputs

Predictor	Pooled r	Benue r (n)	Ekiti r (n)	Lagos r (n)	Interaction p
Total cost / ha	0.53	0.66* (15)	0.13 (24)	−0.38 (19)	0.028
N applied (kg/ha)	0.45	0.11 (16)	−0.41 (24)	−0.21 (19)	0.490

** $p < 0.05$ within state. Interaction p is the F-test p-value for a state × predictor interaction term added to a pooled OLS model of yield on state and the predictor.*

The practical case for high-resolution, regionally disaggregated measurement is direct. The three states in this pilot are not minor variants of a single Nigerian farming system; they sit in different agroecological zones and differ correspondingly in how farming is practiced. Benue lies in the Southern Guinea Savanna, where maize is grown extensively on comparatively large, lower-input plots: average total production cost among our Benue farmers was ₦164,833/ha (US$121/ha) against an average yield of 1.07 t/ha. Ekiti, in the Derived Savanna transition zone, showed markedly higher yields (4.31 t/ha) at only moderately higher average cost (₦208,786/ha, US$154/ha). Lagos, in the Rainforest/Coastal zone and directly adjacent to Nigeria's largest urban market, showed both the highest average investment per hectare (₦989,830/ha, US$729/ha, roughly six times Benue's) and the highest yield (5.72 t/ha), a pattern consistent with proximity to a major urban demand center supporting more input-intensive, higher-value production. A single national, or even single-region, estimate would average over close to a sixfold range in per-hectare investment and a more than fivefold range in yield, obscuring the underlying farming systems rather than describing any of them. Variety choice illustrates the same regional confound at a finer grain: all ten farmers in this sample who grow the traditional Dentcorn variety are in Benue; the activity log shows they paid a real cost for that seed (₦7,500–₦10,000 per farmer, plus a separate seed-treatment cost), so its regional concentration reflects an agroecological and varietal preference specific to Benue rather than free or recycled seed use. Region, agroecological zone, market access, and even seed choice are consequently bound together in this pilot in a way that a study designed or reported only at the national or single-state level would not be able to separate.

Figure 7. Yield versus Total Cost per Hectare, by State

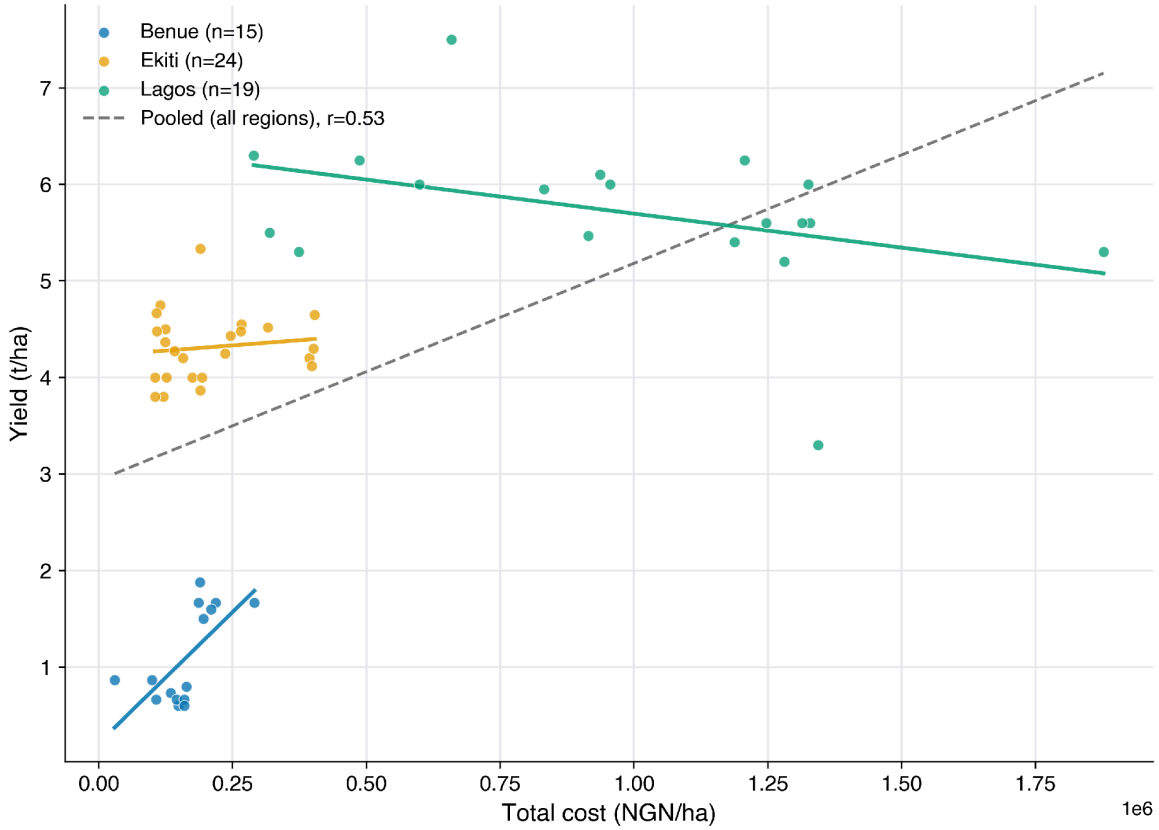


Figure 7. Yield versus total cost per hectare, colored by state. The pooled trend line (dashed grey) is positive; within-state trend lines (solid, colored) diverge, flat or negative in Lagos and Ekiti and positive only in Benue.

5.2. Implications for Agricultural Economic Measurement

This pilot was conducted among Nigerian smallholders because that is where the field-agent network used to validate the instrument already existed, but the measurement problem FETA addresses, farm income and profitability statistics built on infrequent and unverifiable recall, is not specific to smallholder or developing-country agriculture. It maps directly onto several of this conference's stated areas of interest. On farm income and profitability, Section 4.1 shows that, even within the same state, individual farmers' input spending diverges from their peer average considerably more than their outcomes do, direct evidence of farm-level heterogeneity in cost structure that a single aggregate, farm-blind statistic would fail to capture. Whether farmers' own recollection of season cost, revenue, and profit matches this logged detail is a separate question, one being tested directly in an endline survey now under way (Section 6). On labor input measurement, because family and hired labor hours are logged per activity rather than recalled in aggregate, the instrument produces a labor-hours time series anchored to specific field operations, a level of granularity that aggregate labor-cost recall cannot provide. On productivity

measurement using field-level data, the DAP-indexed activity timeline (Figure 2) constitutes, structurally, a field-level micro-panel: it records precisely which inputs were applied, when, and in what quantity, on a specific plot, providing the raw material for field-level productivity measurement rather than a single season-level yield figure. On quality-adjusted intermediate inputs, because each fertilizer application is logged as a distinct product and quantity rather than folded into a season total, the instrument supports converting input-cost data into nutrient-content-adjusted measures, as we do in Section 4.3 rather than relying on cost alone as a proxy for input quantity.

Toward lower-cost data collection: voice-based logging and automated transcription

A natural objection to this design concerns cost. A weekly phone call to every farmer, administered by a trained field agent who manually transcribes the conversation into a structured log, is considerably more enumerator-time-intensive per farm than a single annual recall survey, and does not scale economically to the farm populations relevant to U.S. agricultural statistics. We consider the same event-timeline structure well suited to a substantially lighter-weight, technology-mediated implementation of the same underlying design. Rather than a live phone interview, a farmer could record a short voice message, over WhatsApp or a similar messaging channel already in wide use in both smallholder and U.S. farm contexts, describing that week's activities in their own words, at a time of their own choosing, without requiring live agent time. Automatic speech-to-text transcription, followed by a structured-extraction step using a large language model, could parse that message into the same dated, categorized activity log analyzed in this paper: what was done, what was applied, and at what quantity and cost, for a given week. Under this design, a field agent's role would shift from data collector to exception handler, reviewing only the entries an automated pipeline flags as ambiguous or implausible, the same kind of plausibility checks described in Section 3.3, rather than transcribing every farmer's report manually. We regard this as the most direct path from the pilot reported here to an implementation that could operate, at low marginal cost, across a sample of farms at U.S. scale, converting a research-grade instrument into routine, low-cost, high-frequency measurement infrastructure rather than a one-time study.

This proposal is not merely a matter of technical feasibility; it also depends on whether the relevant stakeholders are willing to participate in this kind of digital, crowdsourced data sharing. Companion survey work by the present authors provides direct evidence on this question across the three stakeholder groups a scaled deployment would involve. Among farmers, a companion survey finds high willingness to share and high perceived value for crowdsourced agricultural information across a range of data types, tempered by only modest willingness to pay for a hypothetical platform, underscoring that participation depends on trust and reciprocal value rather than access to digital tools alone (Abramson and Olowogbon 2026b). Among agricultural advisors (n = 27, across 11 countries), 95.8 percent expressed willingness to share advisory data on a digital platform, and advisors reported already performing a decentralized, informal version of this kind of data intermediation through fragmented tools such as WhatsApp groups and offline spreadsheets (Abramson and Olowogbon 2026a). Among agricultural researchers (n = 70), 98.5 percent expressed willingness to share data on a digital platform, notwithstanding anticipated barriers related to connectivity, digital literacy, and data cost (Abramson and Olowogbon 2026c). Taken together, this evidence suggests that the binding constraint on a lower-cost, crowdsourced version of

FETA is unlikely to be stakeholder willingness; it is more likely to be the data-curation and validation infrastructure needed to make crowdsourced submissions as trustworthy as the field-agent-administered log analyzed in this paper, precisely the intermediation role advisors and researchers, per the same companion surveys, already report performing informally.

Implications for a U.S. deployment

Two further design questions are specific to a U.S. deployment and merit direct discussion, since this conference's audience is well positioned to evaluate them. The first concerns how an instrument such as FETA would relate to data USDA already collects, rather than duplicate it. The National Agricultural Statistics Service's weekly Crop Progress and Condition reports provide a state-level, once-weekly snapshot of crop development stage and the percentage of a state's crop rated in each of five ordinal condition categories (very poor to excellent), aggregated from a network of reporters (USDA NASS, 2025); ARMS provides a detailed but low-frequency, retrospective picture of costs, practices, and income at the farm level (USDA ERS, n.d.). Neither is designed to explain, within the season, why a condition rating changed, which specific input decisions preceded a yield shortfall in a given county, or how a weather shock in a particular week translated into farm-level management responses. A FETA-style weekly activity log, deployed on even a modest subsample of an existing NASS or ARMS sampling frame, could serve as precisely that connective layer: a farm-level, event-dated record against which a Crop Progress anomaly or an unusual signal in county-level yield forecasts could be cross-referenced within the same season rather than the following year. This relationship is close in spirit to the way the LSMS-supported COVID-19 high-frequency phone survey program was designed as a rapid complement to existing household surveys rather than a replacement for them (Gourlay et al. 2021), and we regard the same relationship, FETA as a high-frequency, within-season ground-truth layer alongside USDA's existing survey infrastructure rather than in place of it, as the appropriate framing for a U.S. pilot.

The second question concerns generalizability. Every result in this paper carries the caveat that the pilot sample is a convenience sample rather than a probability sample (Section 3.2), so none of the findings support a population-level claim about Nigerian maize farming, let alone a claim that would transfer to the United States. This limitation has a direct remedy in a U.S. deployment. Rather than recruiting a convenience sample of willing farmers through a pre-existing field-agent network, as in this pilot, FETA could be layered onto an existing probability-based sampling frame, for example a randomly selected, stratified subsample of an ARMS or NASS list or area frame, stratified by region, farm size, and commodity. Under such a design, the leave-one-out peer-benchmarking and deviation analysis in Section 4.1 would generalize to a defined target population, such as corn growers in a specified crop-reporting district, rather than to whoever happened to be reachable through an existing network. A randomized design would also address this paper's other central limitation, the region–enumerator confound documented in Section 5.1. Randomly assigning enumerators, or, following the previous subsection, alternative data-collection channels such as a live call versus a farmer-recorded voice message, across strata, rather than assigning each enumerator to a single geographic area as in this pilot, would allow a U.S. deployment to separately identify region effects, enumerator effects, and channel effects, and would

allow USDA or a research partner to determine which lower-cost channel preserves data quality for which subpopulations before committing to one at national scale.

The exploratory yield results in Sections 4.3 and 5.1 should be read with several limitations in mind: a single growing season, a convenience sample of 64 farms drawn from a pre-existing field-agent network rather than a probability sample, and small per-state cell sizes (15–24 farms). Input-level correlations in particular (cost, applied nutrients) should be read as descriptive rather than causal, given the regional pattern documented in Section 5.1. The peer-deviation results in Section 4.1 rest on firmer ground, since they constitute direct descriptive statistics rather than model-based inferences, although the pooled comparisons in particular should be interpreted alongside that same regional pattern rather than in isolation.

6. Next Steps and Conclusion

Recall accuracy, that is, how closely a farmer's own memory of season cost, revenue, and profit matches the logged detail, is being tested directly by an endline survey fielded as this baseline round closes. The survey asks each farmer for a blind estimate of total production cost, total revenue, and profit for their reference plot before any computed figure is shown. The enumerator then presents the MicroLog-calculated figures for the same categories, including inputs, labor, equipment, transport and storage, other costs, total cost, total revenue, and profit, alongside the farmer's own estimate, and asks which figure the farmer regards as more accurate and why. This design will allow a follow-on paper to report an actual recall-accuracy estimate, which the present paper does not attempt.

One design refinement worth considering for a larger follow-on deployment is a crossed assignment of field agents across geographic areas, with multiple agents per region and each agent covering more than one region, so that a future wave can formally separate region and agent effects in the yield model. We regard this as a robustness check rather than as an indication that the regional pattern reported here is in doubt. We are planning a second deployment with this crossed design, a larger sample, and a multi-season panel structure, which would additionally permit within-farm, across-season comparisons that are not possible with a single-season cross-section.

More broadly, this pilot suggests that a lightweight, weekly, event-timeline instrument, in place of a single post-harvest recall interview, can be deployed at modest cost, can surface farm-level variation in practices and outcomes that a recall-only, single-aggregate design structurally cannot, and can yield a standardized panel suitable for real-time peer benchmarking. Phone-based data collection was an appropriate starting point for validating the instrument, but we do not regard it as the appropriate long-run design; the voice-message-plus-large-language-model pipeline described in Section 5.2 is, in our assessment, what would make this approach viable at U.S. scale rather than as a boutique research instrument. We believe the same underlying design logic, logging events as they occur, standardizing per unit of land, and benchmarking against a leave-one-out peer group, is directly portable to U.S. farm income, labor-input, and field-level productivity measurement, and we would welcome the opportunity to discuss a U.S. pilot, built around that lower-cost data-collection pipeline, with conference participants working in these areas.

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.

Abay, K. A. (2020). Measurement errors in agricultural data and their implications on marginal returns to modern agricultural inputs. *Agricultural Economics, 51*(3), 323–341. https://doi.org/10.1111/agec.12557

Abramson, A., & Olowogbon, T. S. (2026a). Agricultural advisors as data intermediaries: Advisory practices, knowledge exchange, and crowdsourcing readiness. Manuscript under review.

Abramson, A., & Olowogbon, T. S. (2026b). Farmer participation in agricultural crowdsourcing: Perceived usefulness, reciprocal value and willingness to share. Manuscript under review.

Abramson, A., & Olowogbon, T. S. (2026c). From research outputs to field intelligence: Agricultural researchers, crowdsourcing, and the missing feedback loop. Manuscript under review.

Allcott, H. (2011). Social norms and energy conservation. Journal of Public Economics, 95(9–10), 1082–1095.

Arthi, V., Beegle, K., De Weerdt, J., & Palacios-López, A. (2018). Not your average job: Measuring farm labor in Tanzania. Journal of Development Economics, 130, 160–172.

Auta, S. J., & Dafwang, I. I. (2010). The agricultural development projects (ADPs) in Nigeria: status and policy implications. Research Journal of Agriculture and Biological Sciences, 6, 138–143.

BenYishay, A., & Mobarak, A. M. (2019). Social learning and incentives for experimentation and communication. Review of Economic Studies, 86(3), 976–1009.

Carletto, C., Dillon, A., & Zezza, A. (2021). Agricultural data collection to minimize measurement error and maximize coverage. In C. B. Barrett, M. R. Carter, & J.-P. Chavas (Eds.), *Handbook of agricultural economics* (Vol. 5, pp. 4407–4480). Elsevier. https://doi.org/10.1016/bs.hesagr.2021.10.008

Carletto, C., Gourlay, S., & Winters, P. (2015). From guesstimates to GPStimates: Land area measurement and implications for agricultural analysis. Journal of African Economies, 24(5), 593–628.

Chiaka, J. C., Zhen, L., Yunfeng, H., Xiao, Y., Muhirwa, F., & Lang, T. (2022). Smallholder farmers' contribution to food production in Nigeria. Frontiers in Nutrition, 9, 916678.

Conley, T. G., & Udry, C. R. (2010). Learning about a new technology: Pineapple in Ghana. American Economic Review, 100(1), 35–69.

Deininger, K., Carletto, C., Savastano, S., & Muwonge, J. (2012). Can diaries help in improving agricultural production statistics? Evidence from Uganda. *Journal of Development Economics, 98*(1), 42–50. https://doi.org/10.1016/j.jdeveco.2011.05.007

Fafchamps, M., McKenzie, D., Quinn, S., & Woodruff, C. (2012). Using PDA consistency checks to increase the precision of profits and sales measurement in panels. *Journal of Development Economics, 98*(1), 51–57. https://doi.org/10.1016/j.jdeveco.2010.06.004

Gaddis, I., Oseni, G., Palacios-López, A., & Pieters, J. (2021). Measuring farm labor: Survey experimental evidence from Ghana. *The World Bank Economic Review, 35*(3), 604–634. https://doi.org/10.1093/wber/lhaa012

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

Gourlay, S., Kilic, T., Wollburg, P., & Zezza, A. (2021). Viewpoint: High-frequency phone surveys on COVID-19: Good practices, open questions. Food Policy, 105, 102153.

Inegbedion, H., Obadiaru, E., Obasaju, B., Asaleye, A., & Lawal, A. (2019). Financing agriculture in Nigeria through agricultural extension services of Agricultural Development Programmes (ADPs). F1000Research, 7, 1833.

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K.-M., Gerber, J. S., Reddy, V. R., & Kim, S.-H. (2016). Random forests for global and regional crop yield predictions. PLOS ONE, 11(6), e0156571.

Kondylis, F., & Jones, M. (2018). Does feedback matter? Evidence from agricultural services. Journal of Development Economics, 131.

Liverpool-Tasie, L. S. O., Omonona, B. T., Sanou, A., & Ogunleye, W. O. (2017). Is increasing inorganic fertilizer use for maize production in SSA a profitable proposition? Evidence from Nigeria. Food Policy, 67, 41–51.

Nigerian Meteorological Agency (NiMet). (2023). Nigeria Climate Review Bulletin 2023 and 2024 Seasonal Climate Prediction. Abuja: NiMet.

Rodriguez-Segura, D., & Schueler, B. E. (2023). Assessors influence results: Evidence on enumerator effects and educational impact evaluations. Journal of Development Economics, 163, 103109.

Sheahan, M., & Barrett, C. B. (2017). Ten striking facts about agricultural input use in Sub-Saharan Africa. Food Policy, 67, 12–25.

Takeshima, H., & Lawal, A. (2018). Overview of the evolution of agricultural mechanization in Nigeria (IFPRI Discussion Paper 01750). International Food Policy Research Institute.

USDA ERS — U.S. Department of Agriculture, Economic Research Service. (n.d.). What is the Agricultural Resource Management Survey (ARMS)? Retrieved August 2026, from https://www.ers.usda.gov/arms

USDA NASS — U.S. Department of Agriculture, National Agricultural Statistics Service. (2025). Crop Progress and Condition — Guide to NASS Surveys. Page last modified November 19, 2025. Retrieved August 2026, from https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/Crop_Progress_and_Condition/index.php

Wollburg, P., Tiberti, M., & Zezza, A. (2021). Recall length and measurement error in agricultural surveys. Food Policy, 100.