IFPRI Discussion Paper 02420

June 2026

When Quality (Doesn't) Pay

Evidence from Two Experiments in Uganda

Bjorn Van Campenhout
Richard Ariong
Sarah Wairimu Kariuki
Jordan Chamberlin

Markets, Trade, and Institutions Unit

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI), a CGIAR Research Center established in 1975, provides research-based policy solutions to sustainably reduce poverty and end hunger and malnutrition. IFPRI's strategic research aims to foster a climate-resilient and sustainable food supply; promote healthy diets and nutrition for all; build inclusive and efficient markets, trade systems, and food industries; transform agricultural and rural economies; and strengthen institutions and governance. Gender is integrated in all the Institute's work. Partnerships, communications, capacity strengthening, and data and knowledge management are essential components to translate IFPRI's research from action to impact. The Institute's regional and country programs play a critical role in responding to demand for food policy research and in delivering holistic support for country-led development. IFPRI collaborates with partners around the world.

AUTHORS

Bjorn Van Campenhout (b.vancampenhout@cgiar.org) is a Senior Research Fellow in the Markets, Trade, and Institutions (MTI) Unit of the International Food Policy Research Institute (IFPRI), and an Associate Research Fellow at LICOS, KU Leuven, Belgium.

Richard Ariong (r.ariong@cgiar.org) is a Research Analyst in IFPRI's MTI Unit, Kampala, Uganda.

Sarah Wairimu Kariuki (s.kariuki@cgiar.org) is a Markets and Value Chain Specialist at the International Maize and Wheat Improvement Center (CIMMYT), Nairobi, Kenya.

Jordan Chamberlin (j.chamberlin@cgiar.org) is a Senior Economist in CIMMYT's Socioeconomics Program, Nairobi, Kenya

Notices

[1] IFPRI Discussion Papers contain preliminary material and research results and are circulated in order to stimulate discussion and critical comment. They have not been subject to a formal external review via IFPRI's Publications Review Committee. Any opinions stated herein are those of the author(s) and are not necessarily representative of or endorsed by IFPRI.
[2] The boundaries and names shown and the designations used on the map(s) herein do not imply official endorsement or acceptance by the International Food Policy Research Institute (IFPRI) or its partners and contributors.
[3] Copyright remains with the authors. The authors are free to proceed, without further IFPRI permission, to publish this paper, or any revised version of it, in outlets such as journals, books, and other publications.

Abstract

Quality in agri-food supply chains is often unobservable at first sale and early aggregation limits traceability, weakening incentives for quality provision. We study whether making milk quality visible and traceable creates a market for quality in Uganda’s dairy sector. Increasing observability reduces adulteration and improves quality, but no premium emerges. In a follow-up experiment, we introduce trader quality premiums. This increases quality when binding, yet informed intermediaries capture the gains and farm-gate prices do not rise. Observability is necessary but insufficient: without downstream demand for quality and pass-through by intermediaries, incentives for quality upgrading remain weak.

Keywords: Quality measurement, dairy value chains, enforcement, quality premiums, Uganda

Acknowledgments

We would like to thank Denis Athua and Benon Byarugaba of the Ministry of Agriculture Animal Industries and Fisheries (MAAIF). We also would like to thank Leocardia Nabwire, Wilberforce Walukano and Marc Charles Wanume for field support. We are grateful to Alan de Brauw and Xavier Giné for valuable comments on earlier versions of this manuscript. Funding was received from the CGIAR Rethinking Markets Research Initiatives and the CGIAR Better Diets and Nutrition Science Program. We would like to thank all funders who supported this research through their contributions to the CGIAR Trust Fund.

This research received clearance from Makerere University's School of Social Sciences Research Ethics Committee (MAKSSREC-10.2022.594/AR) as well as from IFPRI IRB (DSGD-22-1057). The research was also registered at the Ugandan National Commission for Science and Technology (SS1520ES). Corresponding pre-analysis plans were registered at the AEA RCT Registry as AEARCTR-0010262 (main experiment) and AEARCTR-0016750 (follow-up).

This paper was prepared with the assistance of Claude Code (Anthropic), a large language model. The AI tool was used for code auditing and debugging of R analysis scripts, spelling and grammar checking of manuscript text, and LaTeX/LyX formatting. All analytical decisions, experimental design, data collection, and interpretation of results were made by the authors. The authors take full responsibility for the content of this paper.

1 Introduction

The quality of commodities transacted within value chains, and the preservation of that quality throughout the chain, is central to value chain development (Verhoogen, 2023). Higher-quality inputs improve efficiency by yielding more output per unit of input and reducing waste (Kugler and Verhoogen, 2012). Maintaining quality during aggregation, storage, transport and processing is also critical for food safety and consumer health (Hoffmann et al., 2023). Beyond these direct effects, consistent quality fosters trust and reputation, supports compliance with domestic and export standards, and facilitates access to higher-value markets (Atkin, Khandelwal, and Osman, 2017; Fieler, Eslava, and Xu, 2018). As a result, the transformation of value chains is often accompanied by substantial quality upgrading, reshaping incentives and market structures (Antràs and Chor, 2013).

For quality to become an important driver of behavior, actors must be properly incentivized to produce and preserve it.[1] In many markets, such incentives take the form of price premiums for higher-quality products, which reward the additional effort or investment required to meet quality standards (Shapiro, 1983). When buyers are willing and able to pay more for quality, producers have stronger incentives to adopt better inputs and practices, while intermediaries are more likely to safeguard quality during aggregation, storage, and transport (Bold et al., 2022). In the absence of such premiums, quality improvements may not be financially viable, and actors may instead prioritize quantity over quality (Bai et al., 2025).

The emergence of quality premiums is often hindered by difficulties in observing and verifying quality at the point of transaction. When quality attributes are not directly visible and are costly to discover, buyers face a classic problem of asymmetric information: without credible verification mechanisms, high- and low-quality products trade at similar prices, resulting in adverse selection equilibria in which low-quality goods drive out high-quality ones (Akerlof, 1970). Even in settings where quality can be observed, the absence of systematic verification at the point of sale can prevent quality-based pricing from emerging (Bai, 2025).

Aggregation across suppliers introduces a further and conceptually distinct impediment. When buyers observe only a group-level quality signal (as is common in commodity markets where grading, traceability, and individual branding are absent), individual contributions are observed only with noise, and the shared reputation of the pool becomes a common-pool resource (Winfree and McCluskey, 2005). As Tirole (1996) shows, and as the moral-hazard-in-teams logic of Holmström (1982) implies, no producer fully internalizes the returns to their own quality effort, because all suppliers are paid the same pool price; each can free-ride on the group's standing by supplying below-optimal quality, severing the link between individual performance and reward. The practical consequence is that even when buyers wish to pay for quality, the absence of traceability prevents quality-enhancing investments from translating into higher returns for producers, a pattern well-documented empirically across commodity supply chains in developing countries (Bold et al., 2022; Macchiavello et al., 2026; Bai et al., 2025; Minten, Randrianarison, and Swinnen, 2009). Further constraints arise from the coordination and investment required to establish quality-based payment systems in the first place, including grading infrastructure, standards and (third-party) certification, and monitoring capacity, each of which constitutes a public or collective good subject to its own underprovision problem.

The structure of the value chain itself can exacerbate these challenges. When producers and processors are linked only indirectly through layers of traders or informal intermediaries, the likelihood of traceability breaking down increases substantially (Bai et al., 2025; Macchiavello et al., 2026). By contrast, in chains where producers are more directly connected to processors, opportunities for credible testing, traceability, and clearer communication of standards are greater, increasing the scope for effective quality-based incentives to emerge (Bergquist and Dinerstein, 2020; Hansman et al., 2020).

These challenges are particularly salient in Uganda's emerging dairy sub-sector. Over the past decade, the industry has undergone rapid transformation, especially in the southwestern milk shed around the city of Mbarara, where foreign direct investment has spurred the development of modern value chains (Van Campenhout, Minten, and Swinnen, 2021). A dense network of milk cooling and collection centers (MCCs) now connects dairy farmers to a growing cluster of processors. In dairy, quality is central: it determines processing yields for products such as cheese, casein, and infant formula, and maintaining sanitary conditions is essential given the perishability of raw milk. Yet despite this transformation, there is generally no functioning market for quality in Uganda; prices are

---

[1] In the context of commodities such as maize, milk, or coffee, quality refers to the set of characteristics of a product that determine its value, usability, safety, and suitability for particular buyers or end uses. We use "quality" broadly here; in our empirical setting it refers specifically to compositional and adulteration-related attributes measurable by a milk analyzer (fat, solids-not-fat (SNF), added water), as defined in Section 3.

typically fixed per liter, irrespective of compositional quality attributes such as butterfat, solids-not-fat (SNF), or protein content. Part of the explanation may lie in the lack of technological capacity to measure and communicate compositional quality at the point of aggregation. Most MCCs limit testing to freshness and basic adulteration checks, primarily for added water, and farmers rarely have access to test results. Without the ability to credibly observe and attribute quality at the farm level, the informational preconditions for quality-based pricing are simply absent, making it difficult to establish the kind of verification mechanism necessary for quality premiums to emerge and sustain incentives along the chain.

We use this setting to test whether strengthening quality observability can enable the emergence of quality premiums in Uganda's dairy sector. We implement a cluster randomized controlled trial across four districts in the southwestern milk shed around Mbarara, randomly assigning 95 catchment areas (clusters of one or more co-located MCCs, 125 in total) to receive milk analyzer equipment, digital record-keeping infrastructure, and farmer-facing quality information.[2] Following a one-year intervention period, we measure outcomes at three levels. At the MCC level, we examine whether quality testing became embedded in routine operations (both for incoming farmer deliveries and outgoing sales to buyers) and whether this translated into higher average prices and/or the emergence of explicit quality premiums on either side of the market. At the farmer level, we measure prices received and the incidence of quality bonuses. Finally, to obtain objective measures of compositional milk quality independent of self-reports, enumerators conducted supervised testing of all incoming milk deliveries at both treatment and control MCCs during the course of a full day at endline, yielding direct measures of butterfat, solids-not-fat, protein, added water, and density across the full pool of active suppliers.

The intervention substantially increased the visibility, presence, and day-to-day use of quality-testing infrastructure, generating large and consistent increases in actual testing of both incoming and outgoing milk; the probability that an MCC tests incoming deliveries rose by over 50 percentage points relative to the control group. Despite this operational uptake, the intervention did not affect prices or the emergence of quality premiums at any stage of the chain: farmer sales prices, the incidence of quality bonuses, and all pricing outcomes at the MCC level were uniformly unresponsive, and no quality premiums materialized anywhere along the chain. Most importantly, analyzer installation improved milk quality itself (butterfat increased by around 0.11 percentage points, added water fell by nearly a third, about half a percentage point off a control mean of 1.63 percent, and an overall quality index improved by roughly 0.2 standard deviations) while leaving biologically determined components such as solid-non-fat (SNF) and protein unchanged. This pattern of effects, with adulteration falling but production-side attributes unchanged, is consistent with adjustments in handling and integrity rather than deeper shifts in feeding or animal management practices. Together, the results show that the intervention made quality observable and improved milk composition through reduced adulteration, but without corresponding changes in price incentives or widespread upstream behavioral upgrading: the market for quality did not materialize.

The pattern of results (large increases in testing, measurable improvements in milk composition, yet no movement in prices) is consistent with a straightforward mechanism. The acceptance threshold MCCs apply to incoming milk functions as a floor: once quality measurement makes that floor credible and binding, adulteration falls and composition improves, but there is no incentive to invest beyond what is needed to clear it. The enforcement channel disciplines the bottom of the quality distribution without rewarding the top. Notably, the speed and consistency of the quality response suggests that the chain converges relatively quickly to this new equilibrium, one in which measurement has done its work, adulteration is curtailed, and behavior has adjusted. But it is a low-level equilibrium. The gains from quality measurement are real yet partial: value that could be unlocked through further investment in production practices, feeding, or handling remains uncaptured, because the price schedule offers no signal that such investment would be rewarded.

To test whether activating the pricing channel can shift the chain to a higher equilibrium, we conduct a follow-up experiment at 20 MCCs where the measurement infrastructure is already operational, randomly assigning quality premiums to traders delivering milk to these MCCs. Treated traders receive UGX 100 per liter for each 0.1 percentage point of butterfat and SNF above East African standards, paid daily via mobile money. The experiment proceeds in two stages that jointly allow us to separate the contribution of rewards from the contribution of stricter enforcement to quality upgrading. In Stage 1, thresholds are set at the East African standards (3.3 percent fat and 8.5 percent SNF), a standard that is easy to meet particularly during the rainy season when high-quality milk is naturally abundant and the premium tier is relatively accessible. In Stage 2, the fat threshold is raised from 3.3 to 3.9 percent, simulating the scarcity conditions of the dry season under which sourcing high-quality milk requires

---

[2] In dairy market analysis, a milk shed refers to the geographically defined catchment area from which a processing plant, collection center, or urban market sources its raw milk supply.

active effort from upstream suppliers. Together, the two stages trace how quality-linked contracts affect behavior both when the premium is within easy reach and when meeting it demands genuine investment.

The follow-up experiment reveals that injecting explicit quality premiums into the chain does shift trader behavior, but in ways that depend critically on the scarcity of high-quality milk. When quality thresholds are set at standard East African levels and high-quality milk is naturally abundant, treated traders respond primarily on the volume margin (daily deliveries increase by around 126 liters, almost 50 percent above the control mean) without any meaningful improvement in fat content. The results suggest that when quality is abundant and the threshold easily met, traders find it more profitable to expand the volume of qualifying milk than to invest in sourcing higher-quality milk. When the fat threshold is raised and quality becomes scarce, the pattern shifts decisively: traders move from volume expansion to active quality-based sourcing, delivering milk with significantly higher fat content (up by around 0.17 percentage points), with SNF improving as well. The margin of adjustment changes when the margin that is rewarded changes.

Another striking finding concerns price transmission. Despite receiving explicit, exogenously introduced quality premiums, treated traders do *not* pass value upstream to farmers; if anything, farm-gate prices move slightly downward, by around 17 UGX per liter (about 1.9 percent of the control mean of 906 UGX). One plausible mechanism is informational: the bonus was paid privately to traders and farmers were not informed of its existence, so traders have no competitive pressure from suppliers to share the rent. At the same time, the testing infrastructure gives treated traders verifiable quality readings that farmers cannot independently observe, allowing traders to credibly discount milk they claim is below standard and extract more surplus in the bilateral bargain. Whatever the exact channel, the result underscores a central finding of the paper: even when price incentives are injected into the chain, the structure of intermediation prevents them from reaching the farmers whose behavior ultimately determines raw milk quality at the source.

The paper contributes to three strands of literature. The first concerns the informational preconditions for quality-based markets. The theoretical framework motivating the intervention draws on Akerlof (1970), Tirole (1996), and Winfree and McCluskey (2005), all of which predict that the absence of credible quality measurement severs the link between individual effort and observed outcomes, generating pooling equilibria in which quality investment is socially suboptimal. A growing experimental literature explores how strengthening observability in supply chains alters these equilibria. Saenger et al. (2013) show that, in the Vietnamese dairy sector, third-party verification of previously unobservable quality attributes substantially raises production intensity when embedded in a processor-run contract. Rao and Shenoy (2023) document that group-level quality measurement can overcome free-rider problems in Indian dairy cooperatives, though elite resistance to transparency limits their impact. Our study sharpens this literature by isolating the effect of observability from the effect of pricing: despite a large, credibly identified increase in testing intensity, no quality premiums materialized anywhere in the chain. This result establishes that observability is necessary but not sufficient for quality markets to emerge, and that the downstream contractual infrastructure linking observable quality to price differentials constitutes an independently binding constraint. Xu (2024) arrives at a closely related conclusion in Kenyan dairy cooperatives using a Bayesian traceability design: quality monitoring reduces adulteration through enforcement rather than through new price signals. Together, these studies suggest a recurring pattern in which measurement technology disciplines the lower tail of the quality distribution without activating the continuous quality gradient that higher-value markets require.

Second, the paper contributes to the literature on quality incentive design, a question studied across several commodity settings. Saenger et al. (2013) find that both penalties and bonuses improve farmer investment in quality-enhancing inputs, with bonuses generating the stronger response. Macchiavello et al. (2026) exploit the staggered rollout of a quality sustainability program in Colombian coffee and find that eligible farmers produced higher-quality output in response to embedded price premiums and extension services. Treurniet (2021) finds similarly in Indonesian dairy that individual quality incentives improve compositional quality quickly after introduction, though hygienic quality gains erode over time, suggesting that incentive effects may be fragile even when they successfully reach farmers. In a related setting, Park, Yuan, and Zhang (2025) find that training dragon fruit farmers in Vietnam to meet export quality standards substantially increases quality, with effects amplified when accompanied by buyer incentives. Ashraf, Giné, and Karlan (2009), among others, document that the structure of incentive contracts (in particular whether they reward effort or output) shapes the magnitude and margin of response. Our follow-up experiment, which exogenously introduces quality-linked payments to traders, provides direct evidence on how incentive structure interacts with the scarcity of the rewarded attribute: when quality thresholds are set at standard East African levels and high-quality milk is naturally abundant, treated traders expand volume without improving compositional quality; when the threshold is raised and quality becomes scarce, traders shift from volume expansion to active quality-based sourcing. This pattern maps cleanly onto the margin-of-adjustment logic in the contract theory literature (the behavioral response follows the margin that is actually rewarded) and connects to the broader argument in Verhoogen (2023) that quality upgrading is fundamentally a margin-of-response problem shaped by the structure of incentives and input availability jointly.

Third, the paper contributes to the literature on price transmission through intermediaries in agricultural value chains. Casaburi and Reed (2022) uses a randomized trader subsidy in the Sierra Leone cocoa chain and finds substantial pass-through to farmers, delivered either through prices or credit. The broader experimental record on trader pass-through is, however, far from settled. Bergquist and Dinerstein (2020) use randomized cost shocks among maize traders in Kenya and reach the opposite conclusion: only about 22 percent of the cost reduction reaches consumers, and trader conduct is statistically indistinguishable from joint profit maximization, with intermediaries capturing roughly 82 percent of total surplus. The two studies bracket the question this paper addresses: when does value introduced at the intermediary level transmit through the chain, and when is it retained as rent? Our follow-up experiment speaks directly to this question under experimental logic similar to that of the Sierra Leone study, but in a setting where the value being transferred is informational rather than a flat cost reduction. Traders receiving explicit, daily quality-linked bonuses paid via mobile money do not transmit value upstream; farm-gate prices if anything drift slightly downward, consistent with traders capturing the informational rent created by private, verifiable quality readings that farmers cannot observe. The contrast suggests that unconditional subsidies can generate competitive pressure that raises farmer prices where market structure allows, whereas quality-contingent premiums paid privately to traders, combined with asymmetric access to quality information, allow traders to retain the surplus rather than share it with upstream suppliers. This distinction matters for how the literature interprets pass-through evidence and has direct implications for the design of quality-upgrading programs. Bold et al. (2022) document across four field experiments that even when downstream demand for quality exists, the organization of supply chains often prevents gains from reaching upstream producers; our experiment provides clean identification of the intermediation structure, rather than demand-side uncertainty or information frictions, as the relevant mechanism. Minten and Reardon (2008) offer complementary observational evidence that commodity chains in developing countries frequently fail to transmit quality signals upstream, and Fieler, Eslava, and Xu (2018) and Antràs and Chor (2013) show theoretically that the organization of value chains shapes the incentive to invest in quality at each stage.

We contribute experimental evidence that even when price incentives are cleanly injected at the intermediary level, the structure of intermediation can block their transmission to the farmers whose behavior determines raw material quality at the source. Hansman et al. (2020) show in Peruvian fishmeal exports that firms resolve this non-contractibility by vertically integrating their suppliers, since ownership can realign the quantity-quality trade-off that pay-per-unit contracts leave distorted. We study the market-preserving alternative, making quality observable through measurement technology and injecting explicit premiums while leaving ownership and market relationships intact, and find that observability alone reproduces the enforcement effect of integration, disciplining the lower tail of the quality distribution, but not its continuous quality gradient, because individual quality is never individually rewarded.

Finally, the sequential two-experiment design speaks to a methodological contribution for research on value chain upgrading. Much of the existing literature studies either the informational or the pricing dimension of quality incentives, but rarely both in the same supply chain and with clean separation between them. By first varying observability in a large cluster-randomized trial and then varying pricing within an environment where measurement infrastructure is already operational, we can trace the binding constraint at each stage. The result (that measurement and pricing are complementary rather than substitutable, and that injecting incentives at the trader level does not propagate to farmers) points to a low-level equilibrium trap that cannot be escaped by addressing only one dimension of the problem. This sequential logic connects to the broader argument in Verhoogen (2023) that quality upgrading in developing-country value chains requires coordination across multiple margins simultaneously, and that partial interventions addressing only one bottleneck are likely to generate exactly the kind of limited, compliance-margin improvements documented here.

The remainder of the paper is organized as follows. Section 2 describes Uganda's dairy value chain and the institutional constraints that motivate the study. Section 3 presents the design, data, and results of the main cluster randomized controlled trial. Section 4 presents a simple model that rationalizes the findings and provides the basis for the follow-up experiment, which is presented in 5. Section 6 discusses the results and concludes.

2 Context: Dairy Supply Chains in Uganda

Dairy is the most commercially organized livestock subsector in Uganda, contributing an estimated 4.3 percent of national GDP and 17 percent of agricultural GDP.[3] The 2021 National Livestock Census recorded 1.9 million milked cows producing 3.7 billion liters annually, a fivefold increase from the 667.5 million liters recorded in 2008.[4] Over 2.5 million farming households depend on dairy cattle for income, nutrition, and productive labor. Dairy has also emerged as a significant export earner: export revenues rose from USD 131.5 million in 2018 to USD 264.5 million in 2022/23, with Kenya absorbing over 80 percent of shipments.[5] Despite this growth, per capita milk consumption remains approximately 62 liters per year, well below the 200 liters recommended by WHO, suggesting substantial scope for (domestic) market expansion.

The Ugandan dairy subsector's development is closely linked to significant policy reforms. Notably, the privatization of the National Dairy Corporation and the enactment of the Dairy Industry Act of 1998 marked critical turning points for the industry. These reforms also led to the establishment of the Dairy Development Authority (DDA), a statutory body under the Ministry of Agriculture, Animal Industry, and Fisheries (MAAIF).[6] The DDA had a dual mandate to regulate and promote the dairy sector, ensuring compliance with quality standards while supporting farmers and stakeholders through training, improved technologies, and market development initiatives.

The policy changes spurred an influx of foreign direct investment, particularly in Mbarara, a key town in southwestern Uganda often referred to as the country's "southwestern milk shed". This investment fostered the emergence of a cluster of milk processors in the region, enhancing value addition and market access. Additionally, productivity in the sector has been bolstered by the widespread adoption of improved dairy breeds, such as Holstein Friesians and Jersey cows, significantly increasing milk yields and overall sector efficiency (Van Campenhout, Minten, and Swinnen, 2021).

Although dairy value chains in Uganda differ in their organizational form, from fully integrated systems led by large processors to fragmented structures of small cooperatives, a typical chain comprises five main actors, each with a distinct role in milk production, transport, and processing:

1. Dairy farmers: At the upstream end of the chain are farmer households, who are the primary producers of milk. These farmers deliver raw milk daily to milk collection centers (MCCs) either personally or by relying on intermediaries such as small traders or transporters. Our data suggests that the majority of farmers typically have a motorbike and deliver the milk themselves, which is consistent with the organization of supply chains around cooperatives.

2. Transporters and Traders: Transporters collect milk directly from farms and deliver it to MCCs for a fee, thus providing a service. Traders purchase milk from farmers at the farm gate with the intent to sell it at MCCs or to other traders for profit, thus functioning as commercial intermediaries within the chain. In the follow-up study (Section 5) we work specifically with traders.

3. Milk Collection Centers: MCCs are critical nodes in the value chain. Their primary role is to bulk and chill milk, marking the start of the cold chain essential for maintaining milk quality. These centers are strategically distributed across rural areas, facilitating access for farmers, traders and transporters. MCCs vary considerably in organizational form and capacity. Cooperative MCCs are owned by member farmers, which in principle aligns their interests with suppliers but also creates pressure toward uniform pricing and inclusivity that may work against quality differentiation. Privately owned MCCs operate as commercial entities with more flexibility to differentiate on quality, but less institutional incentive to pass margins upstream.

4. Bulk Traders: Once milk is chilled and bulked at MCCs, large traders transport it to processing facilities. This step often involves the use of specialized milk tankers to preserve quality during transit.

5. Processors: Processors, concentrated in or around key towns such as Mbarara, convert raw milk into value-added products with extended shelf lives, such as ultra-heat-treated (UHT) milk, powdered milk, and infant formula. These processors play a pivotal role in integrating Uganda's dairy sector into both domestic and export markets.

---

[3] FAO-MAFAP, "How MAFAP is Helping to Boost Uganda's Dairy and Beef Sectors," 2024.

[4] Uganda Bureau of Statistics, National Livestock Census 2021.

[5] Dairy Development Authority, Annual Report 2022/23.

[6] The DDA was recently absorbed into MAAIF.

This layered structure, in which milk passes through multiple intermediaries before reaching processors, has important implications for quality incentives. Because milk from different farmers is typically pooled at each stage, traceability from individual producer to final buyer is lost early in the chain, weakening the link between individual quality effort and the price received. The intermediation layers also mean that any quality premium generated at the processor level must pass through traders and MCCs before reaching the farmer, creating multiple points at which price signals can be absorbed or distorted.

Milk quality remains a persistent challenge in Uganda's dairy sector. Media reports and anecdotal evidence from industry stakeholders point to widespread adulteration, most visibly the addition of water to increase delivered volume and, less visibly, the skimming of butterfat for separate sale as cream, ghee, or butter, but systematic data on its prevalence in Uganda are scarce. Evidence from neighboring Kenya, where similar value chain structures prevail, suggests that 5 to 10 percent of milk samples at retail show signs of water addition, with higher rates during the dry season when volumes are low and incentives to dilute are strongest (Wanjohi et al., 2020). Because traders in Uganda are paid per liter with no adjustment for composition, the same incentives apply, and both practices pay: watering inflates the volume sold, while skimming extracts the milk's most valuable component without reducing that volume. Adulterated milk reduces processing yields for butter, cheese, and powdered milk, and in severe cases can lead to rejection of entire batches at the processor level, imposing losses on MCCs. Beyond adulteration, compositional quality (butterfat, protein, SNF) varies substantially across farmers depending on breed, feeding practices, and animal health, but these differences are not reflected in prices.

Pricing along the chain reinforces the low-quality equilibrium. Prices at MCCs are typically fixed per liter, with no differentiation based on compositional quality. Payments from MCCs to farmers are made on a biweekly cycle, creating a wedge between the moment of delivery (when quality could in principle be observed) and the moment of payment (when incentives are realized). A similar rigidity prevails downstream: processors typically announce fixed procurement prices for two-week periods, adjusting only in response to acute shortages or seasonal gluts. This structure severs the link between the quality of a specific delivery and the price paid for it, even when testing infrastructure is available at only one node of the chain.

Quality of raw milk is difficult to observe without proper testing equipment. Most milk collection centers in Uganda rely on two low-cost diagnostic tests to assess basic milk quality: the lactometer test and the alcohol test. The lactometer test is a gravity based test that measures milk density using a floating hydrometer and is primarily employed to detect added water, since dilution lowers density, although readings are sensitive to temperature. The alcohol test, typically using 68 to 75 percent ethanol, assesses protein stability as a proxy for freshness and acidity; milk that curdles upon contact with alcohol is rejected as sour. While both tests are inexpensive and easy to administer, neither provides information on compositional attributes such as butterfat, solids-not-fat, or protein, limiting their usefulness for enforcing or rewarding quality beyond basic adulteration and hygiene.

At present, most MCCs track only the quantity of milk delivered by each supplier, typically recorded by hand in a paper notebook. No compositional quality information is captured at the point of delivery. Electronic milk analyzers offer a more comprehensive alternative: they measure butterfat, solids-not-fat, protein, added water, and density in real time from a small milk sample. However, these devices are not widely available at MCCs due to their cost (approximately USD 1,000-1,500 per unit) and the need for regular calibration and maintenance. When paired with a digital record-keeping system, milk analyzers can provide not only point-of-delivery quality measurement but also systematic tracking of quality across deliveries and over time, enabling MCCs to identify patterns, detect persistent adulteration, and build individual quality histories for suppliers. It is precisely this combination of measurement technology and digital monitoring infrastructure that forms the basis of our intervention in the first experiment.

3 Quality Measurement and Digital Monitoring Experiment

3.1 Design

We use a cluster-randomized controlled trial to test whether making quality observable and traceable at the MCC level improves quality outcomes and generates quality-based pricing along the chain. The unit of randomization is the catchment area, defined operationally as a cluster of one or more nearby or co-located MCCs together with the pool of farmers delivering milk to them at the time of the baseline census. Most catchments contain a single MCC, but in some trading centers two to four MCCs operate in the same location and draw from an overlapping pool of farmers; these MCCs were grouped into a single cluster.

Figure 1: Milk analyzer results during testing showing results for an exceptionally high quality milk sample

The design enables us to estimate treatment effects at multiple levels of the value chain. At the MCC level, we assess whether the system changes testing practices, pricing behavior, and the emergence of quality premiums. At the farmer level, we assess whether farmers connected to treatment MCCs experience changes in prices received, quality bonuses, and production practices. Finally, through supervised milk sample collection and testing at endline, we obtain objective measures of milk quality at the transaction level.

3.2 Intervention

We designed a socio-technical innovation bundle targeting Milk Collection Centers (MCCs). This bundle aims to enhance transparency, improve record-keeping, and empower both MCC managers and farmers by making milk quality measurable and visible. The bundle, developed after extensive consultations with stakeholders and implemented together with the DDA, consists of three key components:

1. Milk Analyzer: A central component of the innovation bundle is the installation of milk analyzers at MCCs. These machines assess milk quality based on a set of compositional parameters, such as butterfat content and solids-not-fat (SNF). Another important parameter is the amount of added water, providing a more accurate assessment of what MCC managers usually test using gravity based methods. The testing process is non-destructive and takes less than one minute per sample, enabling rapid and accurate quality evaluation of each delivery.[7] By providing immediate feedback, the milk analyzers help ensure that quality standards are met and maintained. Figure 1 shows a milk analyzer during piloting.[8]

---

[7] Generally milk was tested per (25 or 50 liter) milk can. Sometimes, if multiple cans of the same farmer were delivered, a composite sample of the different cans was tested.

[8] More information on the milk analyzer can be obtained from the manufacturer's website (https://www.essae.com/milk-fat-analyzer-machine.php).

The milk analyzers were delivered with clear Standard Operating Procedures (SOPs), distributed as laminated reference sheets to remain usable in the day-to-day MCC working environment. In addition, two separate training sessions were organized. The first targeted MCC owners and was held as a half-day workshop in a centralized location, with the focus mostly on generating buy-in by pointing out the benefits of measuring and tracking milk quality and on persuading owners to enforce the new measurement and record-keeping system in their MCCs. The second training was geared toward MCC managers and was delivered individually at each MCC, also lasting about half a day. Its focus was more practical: taking representative milk samples, operating the analyzer, and cleaning and maintaining the device.

To support sustained adoption, the Dairy Development Authority (DDA) conducted periodic monitoring visits to treatment MCCs throughout the intervention year. MCC managers could request technical assistance through a dedicated support channel. Equipment was recalibrated at regular intervals, cleaning reagents were supplied, and malfunctioning units were repaired or replaced.

2. ICT-Mediated Record-Keeping System: We developed an Android application for MCC managers to facilitate digital record-keeping. The app was designed to replace the paper notebooks that MCC managers generally use for keeping track of milk deliveries and payments. In addition to recording quantities and prices, the app allows MCC managers to store and monitor quality parameters obtained from the milk analyzer. The app can provide MCC managers with simple reports, such as the average butterfat (weighted by quantities supplied) over a certain period (today, yesterday, last week, last two weeks, and custom date range). Reports by farmer are also possible, such that MCC managers can determine the total sum to be paid to a farmer for milk delivered over a particular time frame, such as in the last 14 days.[9] In addition to the app, we also developed two online portals that can be used to obtain data for different stakeholders: one portal targeted MCC owners, such that they can monitor key parameters in the MCCs they own; a second portal aggregated information from all MCCs, enabling government officials such as the DDA to monitor quality parameters, prices and quantities in real time. This digital system enhances efficiency and transparency, providing both MCC managers and farmers with reliable records that integrate milk quality metrics. A screenshot of the application can be found in Figure 2.[10]

Milk collection centers were provided with Samsung Galaxy Tab A7 Android tablet computers on which the application was pre-installed. Each tablet contained a SIM card with a prepaid data bundle (which was regularly renewed by the project) to enable cloud-based synchronization of records. The application was developed following offline-first design principles, ensuring that all core functionalities remained accessible without an active internet connection and that data were uploaded automatically once connectivity was restored.

3. Increasing Farmer Awareness of Midstream Testing Capacity: To mitigate potential power imbalances that could arise once MCCs adopt milk analyzers and to ensure farmers are equally informed, we implemented a farmer-facing information campaign using posters displayed at MCCs. Designed by a local artist, the posters publicized the new testing capacity and encouraged farmers to request free quality tests for their milk. By making this service visible and widely known, the campaign aimed to equip farmers with accurate information about their product quality and strengthen trust and cooperation between farmers and MCC managers.

Together, these three components form an integrated strategy to make milk quality observable, traceable, and actionable, improving information flows throughout the dairy value chain.

3.3 Sample and Timeline

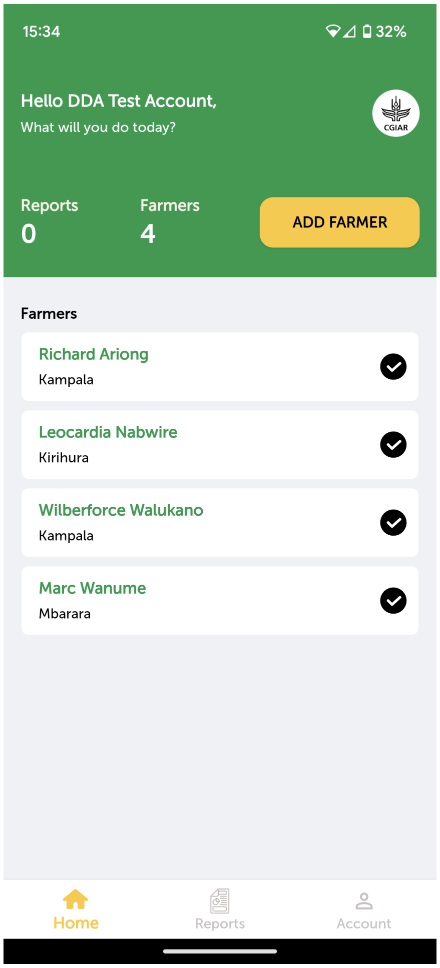
Sample size was determined using a series of power simulations, details of which can be found in the pre-analysis plan.[11] The field experiment was conducted in four districts of Southwestern Uganda: Ntungamo, Mbarara, Kazo,

---

[9] Farmers are typically paid on the 1st and the 15th of the month.

[10] The android app can be downloaded from the Google Play Store: https://play.google.com/store/apps/details?id=com.symatechlabs.ifpri&hl=en

[11] The primary outcome for these calculations was the price of milk, modeled at both the MCC and farmer level. At the MCC level, prices were assumed normally distributed with mean 1000 UGX/liter and standard deviation 50, while farmer-level prices were drawn with the MCC mean and a higher variance (SD = 100) to capture greater dispersion at that level. Processors set fairly stable procurement prices and MCCs adjust tightly around those. The nested nature of the sampling and variance parameters imply an intra-cluster (within MCCs) correlation of approximately $\rho \approx 0.20$. We assumed the intervention would increase MCC prices by 30 UGX/liter (about 3 percent of the assumed baseline, or roughly 0.008 USD/liter at March 2026 exchange rates; a medium to large effect size) and translate into a 40 UGX/liter increase for farmers (about 4 percent, a small to medium effect). Power was calculated at the 5 percent significance level using 1,000 simulations across varying combinations of clusters (100 to 130 MCCs) and farmers per cluster (10 to 40). Results showed that with about 125 MCCs and 20 farmers per MCC (total sample of about 2,500 farmers), power was just above 0.80 for a joint significance test, while power for individual hypotheses was substantially higher, ranging from 0.87 to 0.99. The pre-analysis plan was pre-registered at the AEA RCT registry (https://www.socialscienceregistry.org/trials/10262)



Hello DDA Test Account,
What will you do today?

Reports
0

Farmers
4

Farmers

Richard Ariong
Kampala

Leocardia Nabwire
Kirihura

Wilberforce Walukano
Kampala

Marc Wanume
Mbarara

Figure 2: Dairy record keeping app

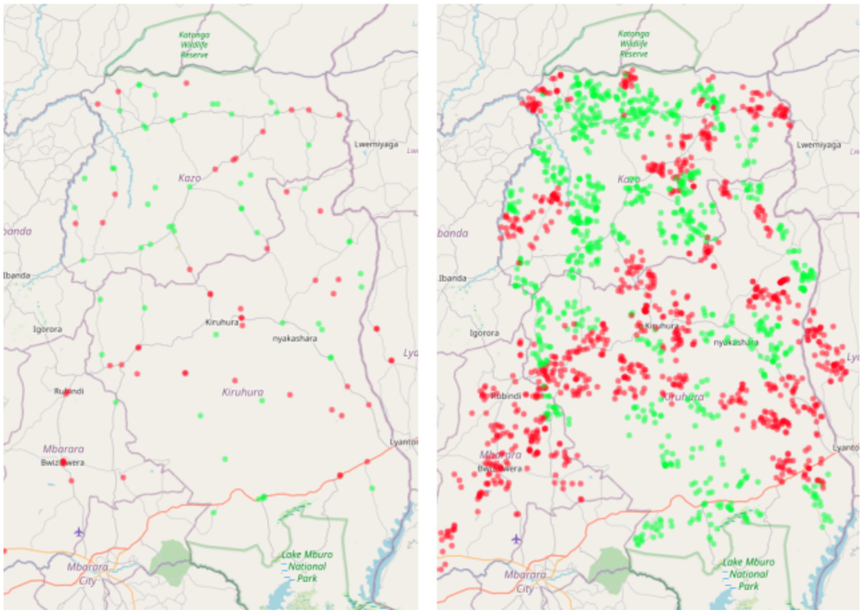


Figure 3: Sampling and Randomization

and Kirihura. The study began with a comprehensive census of all milk collection centers (MCCs) in the region, from which we sampled 125 MCCs. As noted above, co-located MCCs were assigned to the same treatment group, yielding 95 distinct randomization clusters.[12] About half of these clusters were randomly assigned to receive the quality measurement and digital monitoring system, while the other half served as controls. In each of the 125 MCCs, we randomly selected 20 farmers who deliver milk to the center for baseline and endline surveys.

Figure 3 illustrates the randomization in three of the four study districts. The left panel displays the spatial distribution of MCCs and their treatment assignment. The right panel shows the corresponding treatment assignment for farmers connected to those MCCs, reflecting the cluster-randomized design in which all farmers delivering to MCCs in a treated catchment (which may span more than one co-located MCC) are assigned to the same treatment.

Baseline data from both MCCs and associated farmers was collected towards the end of 2022 using in-person surveys. In the second half of 2023, milk analyzers were installed at treatment MCCs and MCC managers were trained on the use of the analyzers and the digital record-keeping system. During the course of one year, together with the DDA we supported treatment MCCs with regular visits to ensure the equipment remained operational, including recalibration after 6 months. In December 2024, endline data was collected. This involved in-person surveys at the MCC level and at the farmer level. In addition, during endline we also measured the quality of incoming milk at each MCC under supervised conditions, yielding 2,518 individual milk samples across the 120 MCCs visited at endline. After endline data collection was complete, milk analyzers were also installed at all control MCCs together with the same training and follow-up protocol used in the treatment arm, both to honour our commitment to participating MCCs and to avoid leaving a quality-measurement gap in the catchments where the intervention had shown positive effects.

3.4 Outcome Variables

We measure outcomes at three levels of the value chain.

At the MCC level, primary outcomes capture the adoption of quality testing infrastructure and the emergence of quality-based pricing. Testing outcomes include whether the MCC uses the milk analyzer to test incoming milk

---

[12] At the design stage, when power calculations were performed and sample size was determined, we did not anticipate the fact that several MCCs share the same location. The implied reduction in sample size due to the clustering raises concerns about statistical power. A minimum detectable effect analysis confirms that the grouping of co-located MCCs does not materially affect statistical power: with baseline controls, the MDE for MCC-level prices is approximately 21 UGX, well below the hypothesized 30 UGX effect, and Fisher randomization inference does not depend on asymptotic cluster counts.

deliveries and whether it tests outgoing milk sold to processors. Pricing outcomes include the average price paid to farmers for milk, the average price received from buyers, and whether the MCC gives or receives quality-based bonuses.

At the farmer level, primary outcomes capture changes in production practices, quality assessment, and market returns. These include an index of improved dairy management practices (oversowing, legume pastures, controlled grazing, pasture conservation, and supplementary feeding), whether the farmer's buyer uses a milk analyzer for quality testing, the quantity-weighted average sales price received, and whether the farmer receives a quality bonus from any buyer.

Finally, to obtain objective measures of milk quality, we collect supervised milk samples at endline. Enumerators equipped with milk analyzers visited each MCC and tested all incoming deliveries over one full day, yielding measurements of butterfat, solids-not-fat, protein, added water, and corrected lactometer reading for each delivery. These visits were unannounced. To limit information spillovers, the testing was conducted over a short period with as many enumerators deployed simultaneously as possible.[13] Secondary outcome families, reported in the appendix, comprise treatment-uptake indicators at both the farmer level and the MCC level, and farmer-level sales-channel choices.

It is important to note that the compositional quality measures obtained through supervised testing at MCCs are farmer-level outcomes, but they need not correspond to the same individuals who appear in the farmer survey. The universe of suppliers observed in the testing data reflects the set of farmers who delivered milk on the specific day of supervised measurement, which may include new entrants who began supplying after analyzers were introduced, as well as exclude farmers who had previously supplied but stopped delivering. Conversely, the survey sample includes baseline suppliers regardless of whether they continued or ceased deliveries by endline. Each data source therefore captures a different margin of adjustment. The MCC testing data reveal changes in the composition and quality of actual incoming supply, incorporating potential selection effects induced by the technology. The survey-based outcomes capture how the intervention affected the behaviors and perceptions of a fixed cohort of farmers.

Outcomes are generally organized in outcome families. To summarize effects across multiple outcomes within a family, we construct Anderson indices (Anderson, 2008), a weighted average of the individual outcomes in which each outcome is first standardized to mean zero and unit standard deviation and then weighted by the inverse of the outcomes' covariance matrix, so that outcomes carrying redundant (highly correlated) information receive less weight. Collapsing a family of related outcomes into a single summary measure is a widely used strategy for mitigating the multiple hypothesis testing problem, as it reduces the number of tests while increasing statistical power. It complements the explicit control of the familywise error rate (FWER) that we describe below in the estimation and inference section (Section 3.5).

3.5 Estimation and Inference

We estimate intent-to-treat (ITT) effects using ordinary least squares with Analysis of Covariance (ANCOVA) specifications that control for baseline values of the outcome where available. At the MCC level, denote milk collection centers by m and let T_m be a treatment indicator equal to one if MCC m lies in a catchment assigned to receive the quality measurement and monitoring system. The MCC-level estimating equation is:

$$y_m = \alpha + \beta T_m + \gamma y_m^b + \varepsilon_m \tag{1}$$

where y_m is the outcome of interest, y_m^b is its baseline value, and β is the parameter of interest; ε_m is an error term that may be correlated within catchment areas.

At the farmer level, let $y_{i,m}$ denote the outcome for farmer i in the catchment area of MCC m:

$$y_{i,m} = \alpha + \beta T_{i,m} + \gamma y_{i,m}^b + \varepsilon_{i,m} \tag{2}$$

where $T_{i,m}$ indicates whether farmer i's catchment was assigned to treatment and $\varepsilon_{i,m}$ is an error term that may be correlated within treatment cluster.

For supervised milk samples collected at endline, we estimate treatment effects using quantity-weighted least squares:

---

[13] This exercise posed logistical challenges, as MCCs opened early (usually at 8:00 am) and were often in remote areas. Enumerators frequently stayed overnight nearby to set up equipment before deliveries and carried generators for use where power was unavailable.

$$q_{j,m} = \alpha + \beta T_m + \varepsilon_{j,m} \tag{3}$$

where $q_{j,m}$ is the quality measure for milk delivery j at MCC m, weighted by the quantity delivered. No baseline control is available for these outcomes as supervised quality testing was conducted only at endline. Also here, we allow $\varepsilon_{j,m}$ to be correlated within catchment areas.

We report point estimates and CR2 cluster-robust standard errors (Imbens and Kolesár, 2016) clustered at the catchment level, which we report for effect-size interpretation and as confidence intervals rather than as our primary test; hypothesis testing relies on the randomization inference described below, whose cluster-level permutation already embeds the design's dependence structure. Because randomization is at the catchment level but sampling is uniform within MCC (twenty farmers per MCC), catchments containing more than one MCC contribute proportionally more observations to the farmer-level regressions. The resulting estimand is a probability-proportional-to-size population-average treatment effect over MCCs and farmers, which is the policy-relevant quantity given that the intervention operates at the MCC level.

Statistical inference is based on Fisher Randomization Inference. Using the OLS coefficient as the test statistic, we evaluate its significance against a permutation distribution generated from 5,000 reassignments of treatment at the MCC catchment level, maintaining the number of treated clusters. To account for multiple testing across outcomes within a family, we apply Westfall-Young stepdown adjustments (Westfall and Young, 1993) using the same permutation distribution, controlling the familywise error rate while exploiting the correlation structure across outcomes; our main results on milk quality and testing uptake remain significant under this correction. Significance stars in all tables are based on the randomization-inference p-values, with CR2 cluster-robust standard errors reported in parentheses for effect-size interpretation.

3.6 Baseline Balance

We pre-registered 10 variables at each level (MCC level and farmer level) to describe the data and demonstrate balance. Results for five variables measured at the MCC level are in Table 1; five more variables are in Appendix Table A1 to conserve space. Table 1 shows that among MCCs in the control group, 63 percent are organized as cooperatives, reflecting how collective institutions continue to play a central role in the organization of Uganda's dairy value chain. The average total storage capacity of MCCs is about 4,000 liters, and 27 percent reported paying a premium for higher quality milk to suppliers.[14] The average MCC has been in operation for just under 10 years, and 58 percent facilitated the supply of acaricides to supplying farmers.

Appendix Table A1 further shows that among MCCs in the control group, the average number of full-time employees is about three, and a typical MCC receives milk from about 56 farmers or traders on an average day during the rainy season. On average, MCCs use 38 percent of their processing or cooling capacity during the dry season, which is indicative of significant seasonality affecting the sector. MCCs report they own 21 milk cans and the vast majority indicate that they provide credit or loans to cooperative members and regular suppliers.

To assess pre-treatment balance across treatment arms, both tables further show treatment-control differences for each baseline variable in the second column. Only one of the individual differences is statistically significant at the 5 percent level (Total capacity of milk tanks). At the bottom of both tables we report results from an omnibus Wald test of joint balance across all covariates (F-statistic and p-values below). These tests do not reject the null of joint equality, indicating that treatment assignment is not systematically predicted by observed baseline characteristics. Overall, the tables suggest that randomization achieved balanced treatment and control groups along key baseline characteristics of MCCs.

Table 2 reports baseline characteristics of dairy farmers by treatment assignment. In the control group (farmers connected to MCCs that did not receive the measurement and monitoring system), household heads are on average 54 years old, with herds of about 62 animals, of which more than 90 percent are improved breeds.[15] Farmers sell

---

[14] While 27 percent of control MCCs report paying a premium for higher-quality milk, this appears to be aspirational rather than operational: MCCs that report paying a premium pay an average of 798 UGX per liter at baseline, statistically indistinguishable from the 790 UGX paid by those that do not (p = 0.63). The 'premium' likely reflects the accept/reject decision rather than continuous quality-differentiated pricing

[15] To calculate herd, we did not simply ask total numbers but asked 6 separate questions: we ask how many local cows; local heifers; and local calves the farmer has, and ask the same 3 categories for improved animals. Farmers were allowed to indicate that they did not know for a particular category, which were treated as missing in our analysis leading to the reduction in the number of observations. As such, the 62 animals include all cattle, so the number of milking animals is substantially smaller. In addition, the distribution is significantly skewed with the median being 47 animals.

Table 1: Balance of Baseline Characteristics Between Treatment and Control Milk Collection Centers (MCCs)

	ctrl	treat	nobs
MCC is cooperative? (1=yes)	0.633	-0.086	124
	(0.486)	(0.090)	
Total Capacity of milk tanks (in liters)	4053.2	1031.1*	124
	(1809.6)	(433.3)	
MCC pays quality premium (1=yes)	0.267	-0.029	123
	(0.446)	(0.082)	
MCC age in years of operation	9.32	0.235	123
	(7.63)	(1.59)	
Facilitates supply of acaricides? (1=yes)	0.583	-0.068	124
	(0.497)	(0.093)	
F-statistic		1.93	
p-value		0.113	

Note: This table reports baseline balance between MCCs assigned to the measurement-system treatment and those in the control group. Column 1 presents control group means with standard deviations in parentheses. Column 2 shows the difference in means (treatment minus control) with robust standard errors in parentheses. Column 3 lists the number of non-missing observations. "Total tank capacity" is measured in liters and "MCC age" in years. The F-statistic and p-value correspond to a joint test of equality of all covariates. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

roughly 60 liters of milk per day during the rainy season, and they spend on average USD 66 per month on chemical purchases, primarily acaricides to fight off tick borne diseases. Further characteristics are in Appendix Table A2 and shows that the average household counts nearly ten members and produces about 70 liters of milk per day during the rainy season (implying they keep about 10 liters for own consumption). Roughly 79 percent of farmers sell most of their milk to an MCC, three-quarters report using only steel containers when transacting (rather than less hygienic plastic jerry cans), and a similarly high share are members of a dairy cooperative.

The farmer-level balance tables (Table2 and Appendix TableA2) also compare pre-treatment characteristics of farmers connected to treatment MCCs with those connected to control MCCs in the second column. Overall, the evidence supports the conclusion that randomization achieved comparable treatment and control groups at the farmer level as well.

3.7 Attrition and Compliance

Meeting sample size targets within the available budget proved challenging at baseline, particularly for the farmer level data because dairy farmers are geographically dispersed. In the end, we surveyed 2,261 farmers across 124 MCCs, corresponding to 87 percent of our planned farmer sample and 99 percent of our planned MCC sample. Because co-located MCCs were grouped together for randomization, these 124 MCCs form 95 distinct catchments (randomization clusters). At the catchment level, 47 were assigned to control (containing 60 MCCs) and 48 to treatment (containing 64 MCCs).

By the time of the endline survey, 4 of the 124 baseline MCCs had attrited: 2 could not be re-contacted and 2 refused to cooperate, leaving us with data for 120 MCCs in 92 catchments. Of the 4 attrited MCCs, one was originally assigned to the control group and 3 to the treatment group. A Fisher exact test does not reject equal attrition rates across treatment arms ($p = 0.62$), though the test has limited power given the small number of attriting units. We further managed to contact 2,139 farmers, of whom 6 refused to be interviewed, leading to an effective sample of 2,133. Further restricting farmers to the subset of farmers connected to an MCC that we were able to interview at endline, we get 2,059. Also at the farmer level, attrition is unrelated to treatment assignment. To examine whether the analysis sample remains representative of the baseline frame, Appendix Table A8 compares baseline characteristics of retained versus attrited farmers. Most differences are small and statistically insignificant, and a joint Wald test fails to reject the null that baseline characteristics jointly predict attrition ($F = 1.55$, $p = 0.16$). Two characteristics differ at the 10 percent level (head age and herd size), with attrited farmers being slightly younger and operating larger herds on average. As a complementary check, Appendix Table A9 re-estimates

Table 2: Baseline Balance of Farmer Characteristics by MCC Treatment Assignment

	ctrl	treat	nobs
Household Head Age (years)	54.47	-0.052	2261
	(12.63)	(0.819)	
Current Total herd size (number)	62.30	4.98	1976
	(54.01)	(6.13)	
Number of improved animals in total herd (share)	0.932	-0.007	1976
	(0.171)	(0.012)	
Liters milk sold per day (on average in the rainy season) (liters)	59.70	5.82	2261
	(56.33)	(6.05)	
Average monthly expense (USD) on chemical purchases	65.55	12.08	904
	(91.10)	(11.48)	
F-statistic		0.268	
p-value		0.927	

Note: This table reports baseline balance between farmers connected to treatment MCCs and those connected to control MCCs. Column 1 presents control group means with standard deviations in parentheses. Column 2 shows the treatment-control difference with CR2 cluster-robust standard errors in parentheses, clustered at the catchment area level. Column 3 gives the number of non-missing observations for each variable. The F-statistic and p-value correspond to a joint test of equality of all covariates. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

the baseline balance regression on the analysis sample. The joint F-test remains comfortably above conventional thresholds (p = 0.61), indicating that selective attrition has not undone baseline balance among survivors.

We also documented several compliance challenges. In some cases, analyzers were moved to control MCCs, and in others, treatment MCCs did not use or retain the machines. At endline, we observed that 14 of the 59 control MCCs had a milk analyzer on site, and 14 of the 61 treatment MCCs did not have a machine readily available. Among the 14 analyzers found in control MCCs, five were devices originally provided through the project, likely reflecting transfers between nearby MCCs. In the treatment group, we identified 47 analyzers across MCCs, but only 37 were operational or could be used for immediate testing at the time of the visit; the remaining units were non-functional or lacked an available operator.

The resulting first-stage relationship between random assignment and actual analyzer availability is 0.37: assignment to treatment raises the probability of having a functioning analyzer from 24 percent in the control group to 61 percent in the treatment group. The ITT estimates should therefore be interpreted as the effect of being assigned to receive the system, averaging over substantial non-compliance in both directions, and are likely attenuated relative to the effect of actual analyzer use. Given this, we supplement our intent-to-treat estimates with local average treatment effect (LATE) estimates that instrument actual analyzer use with random assignment. The fact that we detect significant quality improvements despite this attenuation suggests that the effect on MCCs that actually operate the system is substantially larger than the ITT estimates indicate. The LATE should be interpreted accordingly, as the effect on the sub-population of MCCs whose managers chose to retain and use the analyzer rather than as a generic causal effect of a functioning system. The non-compliance pattern itself is informative: it reveals heterogeneity in how MCC managers value the technology, and the policy-relevant effect at scale will depend on the share of MCCs for which continued operation is worthwhile.

The presence of functioning analyzers at some control MCCs also raises a potential SUTVA concern: if those MCCs acquired project devices through informal transfers from nearby treatment sites, the control group is effectively partially treated, which would attenuate the estimated treatment effect. We return to this concern after presenting the main results, where re-estimating on the clean subsample (dropping the control MCCs with an analyzer on site) yields qualitatively similar and slightly sharper findings; the accompanying LATE estimates, which instrument actual analyzer use with random assignment, provide the natural complement.

3.8 Results

3.8.1 Milk Samples

As mentioned in Section 3.4, to obtain objective measures of compositional milk quality, enumerators conducted supervised testing at both treatment and control MCCs. For one full day at each site, all incoming milk was analyzed using milk analyzers, yielding high-quality measures of key quality parameters: added water, butterfat, solids-not-fat (SNF), protein, and density, measured by the corrected lactometer reading (CLR), a temperature-adjusted density measure reflecting compositional integrity. These variables are also combined into an index of overall quality.

One concern is that supervised testing at endline could induce differential behavior if treatment MCC managers, aware of the analyzers and their parameters, coached farmers or screened deliveries more carefully on the testing day. Several features of the design mitigate this concern. First, testing visits were unannounced and conducted identically at treatment and control MCCs. Second, the quality measures (butterfat, protein, SNF, density) reflect compositional attributes determined primarily by cattle breed, feeding, and handling practices that cannot be altered on short notice. Third, added water, the outcome showing the largest treatment effect, captures adulteration that occurs during transport or at the point of sale, not at the MCC itself. While we cannot fully rule out Hawthorne effects, these considerations suggest they are unlikely to account for the observed patterns.

Table 3 reports the results obtained from the analysis of supervised testing data, yielding 2,518 individual milk samples in total (the unit of observation in this table). Average butterfat content of milk samples delivered to control MCCs was 3.88 percent, rising by an estimated 0.11 percentage points (to approximately 4.0 percent) in MCCs equipped with analyzers. SNF shows a similar directional increase, but the estimate is not statistically distinguishable from zero with the current sample size. Protein, which is a component of SNF, shows no detectable change, consistent with the null finding for SNF.

We also observe improvements in adulteration-related measures. Milk delivered to treated MCCs contains significantly less added water, with a reduction of nearly half a percentage point relative to the control mean

Table 3: Milk quality

	ctrl	treat	nobs
Butter fat (%)	3.88	0.114**	2518
	(0.539)	(0.036)	
SNF (%)	8.58	0.081	2518
	(0.490)	(0.050)	
Added Water (%)	1.63	-0.488*	2518
	(3.68)	(0.213)	
Protein (%)	3.16	0.029	2518
	(0.194)	(0.022)	
Density (CLR)	27.93	0.435*	2518
	(2.31)	(0.197)	
Index	-0.103	0.209**	2518
	(0.817)	(0.054)	

Note: Column 1 reports mean milk quality measures for the control group, with standard deviations in parentheses. Column 2 presents the average treatment effect of the MCC-level measurement-system intervention, with robust standard errors in parentheses. Column 3 shows the number of individual milk samples tested, where each sample corresponds to one delivery brought to an MCC on the day of supervised endline testing. The index is an Anderson (2008) index of the five quality components, with added water entering with a negative sign. Standard errors are clustered at the catchment-area level. CLR denotes the Corrected Lactometer Reading, a standard density-based measure of milk composition adjusted for temperature. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Corresponding LATE estimates are reported in Appendix TableA3.

of 1.63 percent. Density, measured by the corrected lactometer reading (CLR), rises by 0.44 lactometer units, consistent with lower dilution and greater compositional integrity.

The composite index combining all five components mirrors these patterns, indicating a significant improvement of 0.21 standard deviations in overall milk quality. Taken together, these results suggest that the introduction of analyzers reduced adulteration and improved butter fat content, but had limited effects on biologically determined components such as SNF and protein. This pattern is consistent with adjustments driven by changes in handling practices (such as reduced skimming or dilution) rather than by deeper shifts in feeding or production management, in line with the comparatively weak farmer-level behavioral responses documented in the next subsection.

The distributional pattern of the added-water variable, illustrated in Figure 4, is particularly informative. While the median delivery shows no detectable added water, approximately 29 percent of deliveries in the control group test positive for some water addition, and 11 percent show more than 5 percent added water. In the treatment group, the incidence of serious adulteration (more than 5 percent added water) falls to 5 percent, roughly half the control rate. The treatment effect thus appears concentrated in the right tail of the adulteration distribution: the intervention disciplines the most egregious adulterators without necessarily changing behavior among suppliers who were already delivering clean milk.

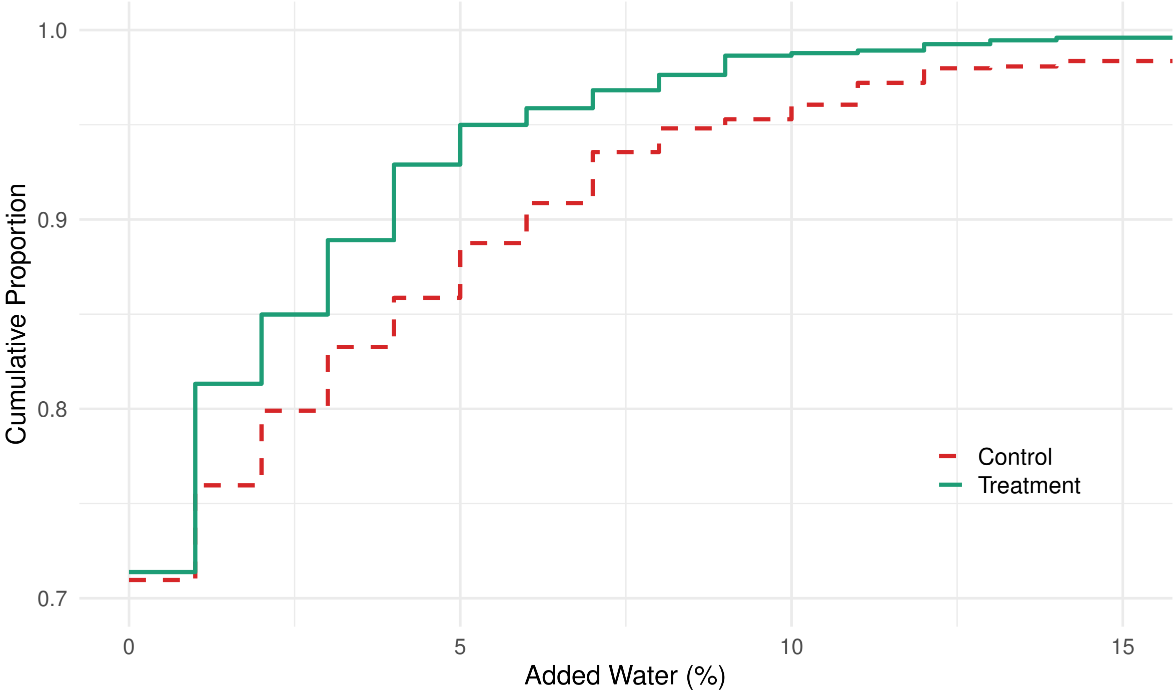


Figure 4: Empirical CDF of added water (%) by treatment status.

The treatment-on-the-treated (LATE) estimates, presented in Appendix Table A3 for completeness, are generally less precisely estimated than the intention-to-treat (ITT) effects. The LATE point estimates are roughly two to three times larger than the corresponding ITT coefficients, reflecting the rescaling by the compliance rate. The composite quality index remains statistically significant at the 5 percent level under 2SLS, and the individual estimates for fat content and density (CLR) are marginally significant, while added water loses statistical significance despite retaining the expected signs. The qualitative conclusions are thus unchanged: actual analyzer use improves milk composition.

3.8.2 Farmer Level

We also define four primary outcomes at the farmer level and combine them into an index to assess the overall impact of the intervention on farmer-level outcomes.

Table 4: Primary Outcomes for Farmers

	ctrl	treat	nobs
Production investment and management (Index)	0.014	-0.020	2054
	(0.562)	(0.039)	
Buyer checked for quality (1=yes)	0.157	0.071	1337
	(0.364)	(0.047)	
Price received for milk sold	1020.1	11.31	1254
	(97.59)	(9.94)	
Get quality premium	0.081	-0.016	1281
	(0.274)	(0.023)	
Index of primary farmer outcomes	-0.036	0.080^{+}	1202
	(0.490)	(0.047)	

Note: Column 1 reports the mean of each outcome for control farmers, with standard deviations in parentheses. Column 2 shows the estimated treatment effect of the measurement-system intervention, with robust standard errors in parentheses. Column 3 lists the number of farmers with non-missing observations. Prices are in Ugandan shillings (UGX). The "Index of primary farmer outcomes" is an Anderson (2008) index of the variables listed above. Standard errors are clustered at the catchment-area level. All estimates are intent-to-treat (ITT). Corresponding LATE estimates are reported in Appendix TableA4. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ (significance stars based on randomization-inference p-values).

The first outcome is a composite in itself, and measures production investment and management practices that are expected to improve milk quality. We construct an Anderson (2008) index from farmer reports on six recommended practices undertaken in the past year: (i) over-sowing new fields with improved forage grasses such as *Napier*, *Brachiaria*, or *Rhodes* grasses; (ii) using legume pastures such as *Centro*, *Siratro*, or *Desmodium*; (iii) adopting controlled or zero grazing during the last dry season; (iv) practicing pasture conservation through hay, silage, or haylage; and (v) using feed supplements such as maize bran, crop residues (for example banana peelings), or mineral licks. This index summarizes the extent to which farmers adjusted production strategies in response to the intervention.

The second outcome captures whether buyers actively checked milk quality at the point of transaction. Here, we distinguish between different categories of buyers (e.g., MCCs, processor, trader, etc.) and ask if milk was tested using a milk analyzer if the farmer sold milk to this particular buyer in the 7 days prior to the survey. The indicator is true if milk sold in the last 7 days was tested by at least one buyer.

The third outcome is the price received for milk sold. Specifically, we record the price per liter obtained in the farmer's most recent sale during the past seven days, inclusive of any quality premiums. Farmers reported prices by buyer category, and we construct a weighted average across categories using transaction volumes as weights. We also directly ask whether the buyer pays a higher price for higher-quality milk. This question was posed in general terms, without reference to a specific time frame. These outcomes capture whether quality improvements translate into tangible financial benefits for farmers. Note that the price and quality-bonus outcomes are defined only for farmers who sold milk in the seven days preceding the survey; the observation count for these outcomes therefore falls below the full analysis sample of 2,058 (to roughly 1,250-1,340, depending on the outcome). [16]

Table 4 reports treatment effects on our four primary outcomes at the farmer level, as well as the composite index. The table shows ITT estimates from Equation 2 in Section 3.5. The first column reports the control group mean (with standard deviations in parentheses), the second column reports the estimated treatment effect of the measurement-system intervention (with cluster-robust standard errors in parentheses), and the third column reports the number of non-missing observations. In light of the compliance issues reported in Section 3.7, we also present LATE estimates instrumenting actual analyzer availability with randomized treatment assignment (See Appendix

---

[16] Because the act of selling is itself potentially affected by the intervention, Appendix Table A12 reports treatment effects on sales-channel outcomes, including the probability that a farmer sold milk to the MCC. Treated farmers are marginally less likely to have sold to the MCC in the wet and dry seasons (a reduction of roughly seven percentage points in each case, significant at the ten percent level), indicating mild selection into the price-conditional analysis sample; conditional on having sold, however, average prices received are statistically indistinguishable between treated and control farmers in either season.

Table A4).

We do not find evidence that the measurement-system intervention affected farmers' production investments and management practices. The point estimate is small and statistically insignificant. Among control farmers, approximately 16 percent report that their milk was tested using an analyzer. Farmers linked to treatment MCCs report slightly higher testing rates (roughly 7 percentage points more), although this difference is not statistically distinguishable from zero.[17]

We find no evidence that the intervention affected the prices farmers receive for milk. Control farmers report an average price of UGX 1,020 per liter, a level consistent with MCC-level data once accounting for sales to alternative buyers such as traders or direct consumers, who typically pay slightly less. Prices received by farmers in the treatment group are statistically indistinguishable from those of control farmers. Similarly, the likelihood that a farmer reports receiving an explicit quality premium does not differ between groups, mirroring the null effects observed for prices. These results point to no detectable impact of the intervention on the price dimension of farmers' marketing outcomes.

While no individual outcome reaches conventional significance, the composite index of primary farmer outcomes is marginally significant ($p < 0.10$), driven primarily by the increase in buyer quality testing. Results for the LATE analysis in Appendix TableA4 are very similar.

3.8.3 MCC Level

We pre-registered six primary outcomes at the level of the MCC. All six are hypothesized to move in a positive direction in response to our intervention designed to make milk quality more transparent and traceable. To capture the aggregate effect across these outcomes (and simultaneously consider multiple hypothesis testing concerns), we again construct a summary index following Anderson (2008).

Our first two outcomes measure the extent to which quality testing becomes embedded in routine operations at the MCC. Specifically, we record whether MCCs used a milk analyzer to measure butter fat and solids-not-fat content of incoming milk samples in the last 7 days. In addition, we asked whether MCCs test outgoing milk deliveries destined for buyers. This measure is based on more detailed sales transaction data, and is true if for any buyer (e.g., processor, other trader, etc) butter fat or SNF was tested using a milk analyzer. Together, these indicators capture the degree to which the analyzers are used as intended, and whether testing shapes quality assurance both upstream and downstream.

The next two outcomes reflect how increased transparency affects prices. We measure the average price at which MCCs purchase milk from suppliers during the seven days preceding the survey. For onward sales by MCCs, we ask about the average price MCCs received for sales to various buyers during the last transaction in the previous week, and take a (weighted) average of these. These outcomes allow us to assess whether improved quality monitoring translates into higher farm-gate prices and whether MCCs are able to cash in on quality improvements through higher sales prices.

Finally, two outcomes directly capture whether price differentiation by quality emerges. We ask whether the MCC pays explicit quality premiums to its suppliers, and whether downstream buyers pay a premium for higher-quality milk. Together, these indicators provide evidence on whether the intervention helped overcome the central coordination problem in quality upgrading: aligning incentives on both sides of the market so that producers and intermediaries are rewarded for investing in quality.

Results are in Table 5. The results show that many MCCs saw the merits of milk analyzers. In particular, MCCs in the treatment group were substantially more likely to test incoming deliveries: the probability of testing rose by 53 percentage points relative to a control group mean of 29 percent, a large and highly significant increase. We also observe a large and significant increase in the probability of testing outgoing deliveries to buyers. These results suggest that the analyzers were effectively used for monitoring both milk supplied by farmers and deliveries to downstream buyers.[18]

Turning to prices, we find no evidence that the intervention affected transaction terms at MCCs. In the control group, the average farm-gate price paid to suppliers was about UGX 1,075 per liter, and we detect no significant

---

[17] Although these effects are generally small and imprecise, they are consistent with qualitative reports from MCC managers who stated that analyzers were typically used only in suspected cases of adulteration rather than on every single delivery. Even so, selective but credible testing can still shift behavior: the possibility of being tested may be sufficient to induce farmers to reduce bad practices or upgrade milk quality.

[18] These self-reported outcomes may raise concerns about experimenter bias. However, as shown in Appendix Table A7, we obtain similar results when using direct enumerator observations of analyzer use.

Table 5: Primary Outcomes for Milk Collection Centers (MCCs)

	ctrl	treat	nobs
Tested incoming milk using MA (1=yes)	0.288	0.530^{**}	120
	(0.457)	(0.081)	
Testing outgoing milk using MA (1=yes)	0.203	0.288^{**}	120
	(0.406)	(0.089)	
Price at which milk was bought from suppliers (UGX)	1075.0	-15.21	115
	(92.56)	(16.36)	
Price at which milk was sold (UGX)	1199.6	3.44	108
	(106.3)	(19.22)	
Does the MCC pay a quality premium to suppliers?	0.186	-0.032	119
	(0.393)	(0.066)	
Did the buyer pay a quality premium?	0.186	0.031	119
	(0.393)	(0.071)	
Index of primary MCC outcomes	-0.077	0.174^{+}	115
	(0.477)	(0.096)	

Note: Column 1 reports the mean of each outcome for the control group, with standard deviations in parentheses. Column 2 shows the estimated treatment effect of the measurement-system intervention at the MCC level, with robust standard errors in parentheses. Column 3 lists the number of MCCs with non-missing observations. All prices are in Ugandan shillings (UGX). The "Index of primary MCC outcomes" is an Anderson (2008) index of the variables listed above. Corresponding LATE estimates are reported in Appendix TableA5. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

increase in the treatment group. Similarly, the average price MCCs received from buyers was about UGX 1,200 per liter in the control group, with no evidence that the MCC level treatment raised buyer offers. These results suggest that while the intervention influenced testing practices, it did not translate into higher prices on either side of the market.

In addition, about 19 percent of MCCs in the control group reported paying a quality premium to suppliers, and the same share (11 out of 59) reported receiving a premium from buyers. Consistent with the price outcomes, we do not find that making quality more visible in the value chain increased the likelihood of such premiums being paid.

In light of the compliance issues discussed in the previous section, we also estimated Local Average Treatment Effects (LATE), instrumenting the actual availability of a functioning milk analyzer with the randomized treatment assignment (See Appendix Table A5). The conclusions remain largely unchanged. Overall, MCC-level data indicate that the intervention produced a sizable and consistent increase in testing activity, yet we find no evidence that it generated quality-based price differentiation. When we consider the composite index of pre-registered MCC-level outcomes, the data indicate that the measurement-system intervention led to marginal improvements at the MCC level, but this is entirely driven by an increase in MCC milk testing behaviour.

3.8.4 Secondary outcomes

We examine several pre-registered secondary outcomes capturing treatment uptake and compliance at both the MCC and farmer level (Appendix Tables A6 and A7). At the MCC level, the treatment produced large and precisely estimated effects on technology presence and use. Treated MCCs are significantly more likely to display project posters publicizing the analyzers, to have a functioning milk analyzer on site (an increase of roughly 53 percentage points), and to use digital record-keeping tools for quality and quantity data. These results confirm both compliance with treatment assignment and the persistence of the hardware and associated practices at treated sites. The adoption of digital record-keeping is particularly noteworthy, as it addresses the traceability and transparency that are prerequisites for quality-linked transactions.

At the farmer level, we find no evidence that the intervention affected secondary outcomes, including seed usage or understanding of compositional quality. This is consistent with the intervention design: the treatment targeted technology adoption at collection centers, not farmer education. Improved analyzer use at MCCs did not spill over into farmer knowledge or practice, reinforcing the interpretation that the intervention operated through the enforcement channel at the MCC level, without affecting the incentive structure that would lead to changes in farmer behaviour.

3.8.5 Robustness to Control-Group Contamination

As noted in Section 3.7, 14 of the 59 control MCCs had a functioning milk analyzer on site at endline, raising the possibility that informal transfers between nearby MCCs contaminated the control group and attenuated our estimates. To assess the quantitative importance of this concern, we re-estimate the main specifications on the subsample of 106 MCCs (and their 1,814 associated farmers) that remains after dropping these 14 contaminated control MCCs. Full results are reported in Appendix Table A13, which places full-sample and clean-sample treatment effects side by side for the three result levels (milk samples, MCC outcomes, farmer outcomes). The point estimates are qualitatively unchanged and, consistent with attenuation from contamination, somewhat sharper. The composite milk-quality index rises by 0.22 standard deviations on the clean sample, compared to 0.21 on the full sample, and the individual components of the index (fat, added water, CLR, protein) move in the same direction and retain or gain statistical significance. At the MCC level, the probability of testing incoming milk with a functioning analyzer increases by 75 percentage points on the clean sample, compared to 53 in the full sample, while the other MCC-level outcomes (farm-gate and buyer prices, quality premiums) remain statistically indistinguishable from zero. At the farmer level, the composite index of primary outcomes becomes statistically significant at conventional levels (coefficient 0.13, $p = 0.007$), compared to marginal significance in the full sample, driven again by the increase in buyer testing of milk quality. We conclude that the observed contamination does not drive the main findings, and that if anything it has caused us to understate the effect of a functioning quality measurement infrastructure.

4 A Model of Quality Infrastructure and Quality Incentives

This section presents a simple analytical framework to rationalize the findings from the experiment. The model formalizes two channels through which quality infrastructure can affect outcomes: an enforcement channel, in which improved measurement precision makes rejection thresholds credibly binding, and a price channel, in which quality measurement enables quality-differentiated pricing. The main experiment results are consistent with the enforcement channel operating in isolation: quality improves but prices do not differentiate. We then present results from a follow-up experiment that tests the price channel directly by introducing quality-based premiums at the trader-MCC interface, bypassing the processor-MCC pass-through that would be required for a market to emerge.

4.1 Environment and Timing

There is one processor, multiple MCCs indexed by m, and, within each MCC catchment, multiple farmers indexed by i that deliver milk to the MCC. The quantity of milk supplied by farmer i, denoted y_i, is assumed to be fixed in the short run, as it primarily depends on the farmer's herd size, which does not adjust quickly. Milk quality is expressed on a continuum ($q_i \in [0, 1]$) and reflects compositional attributes such as butterfat and SNF.

After observing delivered quantities and a (possibly noisy) signal of quality for each farmer, the MCC decides whether to accept or reject each delivery. Accepted milk is aggregated and sold to the processor, who pays a premium for verifiable compositional quality. The downstream price received by the MCC therefore depends on average quality of accepted milk.

A critical feature of the model is how the MCC observes quality. Without measurement technology, the MCC observes a noisy signal of farmer i's quality:

$$s_i = q_i + \eta_i \tag{4}$$

where $\eta_i \sim N(0, \sigma^2)$ is measurement noise. With only basic tools (lactometer, alcohol test), σ^2 is large: the MCC can detect gross adulteration but cannot precisely assess compositional quality. The introduction of a milk analyzer reduces σ^2 toward zero, enabling precise measurement of butterfat, SNF, protein, and added water.

The acceptance decision is based on the observed signal:

$$a_i = \mathbf{1}\{s_i \geq q_{min}\} \tag{5}$$

rather than true quality q_i. This distinction is central: when σ^2 is high, the acceptance threshold is only weakly enforced because the MCC cannot reliably distinguish high-quality from low-quality milk. Reducing σ^2 makes the threshold bite.

Farmers are connected to the MCC either directly or through intermediaries, which may affect the extent to which downstream quality incentives reach them.

4.2 Production of Quality at the Farm

Each farmer i supplies a fixed quantity of milk y_i in the short run and chooses effort $e_i \geq 0$ that affects the compositional quality of milk:

$$q_i = \theta_i e_i + \varepsilon_i \tag{6}$$

with $\theta_i \geq 0$ capturing how effectively effort is translated into quality, and ε_i a mean-zero shock.

Effort is assumed to be costly, which we model as a quadratic function:

$$c_i(e_i) = \frac{k_i}{2} e_i^2 \tag{7}$$

with $k_i > 0$.

4.3 Prices, Pass-Through, and Intermediation

Let the processor pay the MCC a unit price

$$p_m = \bar{p} + \alpha \, \bar{q}_m \tag{8}$$

where $\bar{q}_m$ is average quality in MCC m and $\alpha \geq 0$ is the per-unit premium slope for verifiable quality. In practice, α depends on whether the MCC can credibly demonstrate quality to the processor. When σ^2 is high, the MCC cannot verify its own quality and has no basis to negotiate quality-differentiated prices; processors, in turn, have little reason to offer quality premiums if upstream infrastructure cannot reliably deliver consistent compositional quality. When σ^2 falls, two things change. First, the MCC acquires verifiable quality information that strengthens its bargaining position: it can credibly demonstrate to the processor that its milk exceeds the market average and can approach competing buyers with evidence. Second, processors may find it worthwhile to offer quality-differentiated contracts because upstream measurement infrastructure makes consistent delivery credible. Whether α actually increases depends on whether these conditions are sufficient to overcome contracting frictions, processor market power, and coordination failures. A central question of this paper is whether equipping MCCs with analyzers is sufficient to activate $\alpha > 0$.

The MCC posts a pricing rule that links observed quality (q_i) to price paid to farmers (p_i). Let the supplier-level price schedule be:

$$p_i = p_0 + \lambda_i \alpha \, q_i \tag{9}$$

where p_0 is a base price, and $0 \leq \lambda_i \leq 1$ is the pass-through rate of the premium to farmer i. Pass-through depends on local contracting frictions, transparency, and the structure of intermediation.

4.4 Farmer Optimization

The MCC may reject milk that does not meet a minimum quality standard. Let $a_i \in \{0, 1\}$ be an acceptance indicator, where $a_i = 1$ if farmer i's observed signal meets a minimum quality threshold q_{min} and $a_i = 0$ otherwise:

$$a_i = \mathbf{1}\left\{s_i \geq q_{min}\right\} \tag{10}$$

Farmer i's realized revenue is then

$$r_i\left(e_i\right) = a_i p_i\left(q_i\right) y_i \tag{11}$$

Because both quality and its observation include noise, acceptance is uncertain from the farmer's perspective. Since $s_i = q_i + \eta_i = \theta_i e_i + \varepsilon_i + \eta_i$, and ε_i and η_i are independent, the acceptance probability can be written as:

$$\pi_i\left(e_i; q_{min}, \sigma^2\right) = \Pr\left(s_i \geq q_{min} | e_i\right) = 1 - \Phi\left(\frac{q_{min} - \theta_i e_i}{\sqrt{\sigma_\varepsilon^2 + \sigma^2}}\right) \tag{12}$$

where Φ is the standard normal CDF and σ_ε^2 is the variance of ε_i. This makes explicit that reducing σ^2 (better measurement) tightens the acceptance probability around the threshold.

The acceptance probability is increasing in effort:

$$\frac{\partial \pi_i\left(e_i; q_{min}, \sigma^2\right)}{\partial e_i} > 0 \tag{13}$$

Farmer i's expected revenue is therefore

$$E\left[r_i\left(e_i\right)\right] = \pi_i\left(e_i; q_{min}, \sigma^2\right)\left[p_0 y_i + \lambda_i \alpha \theta_i e_i y_i\right] \tag{14}$$

The farmer chooses effort to maximize expected net returns:

$$e_i^* = \arg\max_{e_i \geq 0} \left\{\mathbb{E}[r_i(e_i)] - c_i(e_i)\right\} \tag{15}$$

This decision problem highlights two distinct channels through which effort affects expected payoffs:

1. Intensive margin (price effect): When pass-through is positive ($\lambda_i > 0$), higher effort raises quality q_i and therefore the quality-linked component of the price $\lambda_i \alpha q_i$ conditional on acceptance. Holding the acceptance probability fixed, the marginal return to effort comes from a steeper price–quality gradient.
2. Extensive margin (acceptance effect): Even absent pass-through on the price (for example when $\lambda_i = 0$), higher effort increases the probability that milk meets the MCC's standard and is accepted, $\pi_i\left(e_i; q_{min}, \sigma^2\right)$. If rejected milk receives a strictly lower payoff than accepted milk (for example zero or a low salvage value), this acceptance channel generates its own incentive to exert effort.

The first-order condition, which implicitly defines optimal effort is

$$\frac{\mathrm{d}}{\mathrm{d}e_i}\mathbb{E}[r_i(e_i)] = k_i e_i, \tag{16}$$

where the left-hand side collects both intensive and extensive margin effects:

$$\frac{\mathrm{d}}{\mathrm{d}e_i}\mathbb{E}[r_i(e_i)] = \underbrace{\pi_i(e_i; q_{min}, \sigma^2)\lambda_i \alpha \theta_i y_i}_{\text{intensive margin}} + \underbrace{\frac{\partial \pi_i(e_i; q_{min}, \sigma^2)}{\partial e_i}\left[p_0 y_i + \lambda_i \alpha \theta_i e_i y_i\right]}_{\text{extensive margin}}. \tag{17}$$

The model therefore implies that stronger quality-based incentives at the MCC level (higher α), higher pass-through λ_i, more effective technology θ_i, and lower effort cost k_i, all increase optimal effort e_i^*, realized quality $q_i^* = \theta_i e_i^*$, and the likelihood of supplying accepted milk.

4.5 MCC Pricing, Screening, and Outcomes

The MCC observes a (possibly noisy) measure of individual quality and decides whether to accept each farmer's delivery. Aggregate quality at the MCC is then a weighted average of the quality of all accepted milk:

$$\bar{q}_m = \frac{\sum_i a_i y_i q_i}{\sum_i a_i y_i} \tag{18}$$

The MCC chooses a quality standard (acceptance rule) that trades off higher average quality against the loss of volume from rejected milk. A higher weight on quality in the downstream price schedule, α, makes it optimal for the MCC to tighten this standard, rejecting more low-quality milk and thereby directly increasing $\bar{q}_m$, even holding farmer effort fixed.

The MCCs aggregate milk and sell to the processor. Its per-unit margin per farmer is:

$$\underbrace{\bar{p} + \alpha\bar{q}_m}_{\text{revenue from processor}} \quad - \underbrace{(p_0 + \lambda_i \alpha q_i)}_{\text{payment to farmer}} \tag{19}$$

Averaging over farmers, the expected margin per liter becomes:

$$\mu_m = \bar{p} - p_0 + \alpha(1 - \bar{\lambda}_m)\,\bar{q}_m \tag{20}$$

where $\bar{\lambda}_m$ is the (volume-weighted) mean of the individual pass-through rates applied to accepted milk.

An increase in α raises $\bar{q}_m$ both directly (through stricter acceptance), and, at least if $\lambda_i > 0$, through increased effort of the farmer.

4.6 Model predictions

4.6.1 Effect of Improved Measurement Precision

The central comparative static for the main experiment concerns the effect of reducing measurement noise σ^2. Consider the acceptance probability:

$$\pi_i\left(e_i; q_{min}, \sigma^2\right) = 1 - \Phi\left(\frac{q_{min} - \theta_i e_i}{\sqrt{\sigma_\varepsilon^2 + \sigma^2}}\right) \tag{21}$$

Reducing σ^2 (improving measurement precision) has two effects. First, for farmers whose expected quality $\theta_i e_i$ is close to but above the threshold q_{min}, acceptance becomes more certain, reducing the risk of false rejection. Second, and more importantly, for farmers whose expected quality is near or below the threshold, the probability of rejection increases sharply. The threat of rejection, previously weakened by the noise in the MCC's observation, becomes credible.

This tightening of the acceptance probability distribution around the threshold generates an incentive for farmers near the margin to increase effort:

$$\frac{\partial e_i^*}{\partial \sigma^2} < 0 \quad \text{when } \theta_i e_i \text{ is near } q_{min} \tag{22}$$

That is, better measurement (lower σ^2) induces higher effort from farmers who are close to the acceptance threshold. Crucially, this mechanism operates entirely through the extensive margin (acceptance probability) and does not require any price differentiation (α can be zero). The model therefore predicts that measurement technology can improve quality purely through more credible enforcement, without any change in the price schedule.

This is the enforcement channel. Its key empirical signatures are: (i) quality improves, particularly through reduced adulteration (the quality dimension most directly affected by effort near the threshold); (ii) prices do not change (because reducing σ^2 alone was not sufficient to activate $\alpha > 0$); (iii) the quality improvement is concentrated at the lower tail of the distribution (farmers near q_{min} adjust, while those already well above the threshold do not).

The enforcement channel explains the main experiment results: quality improves through more credible rejection threats, but prices do not differentiate. The model, however, also generates predictions about what would happen if quality-based pricing were activated, i.e., if α were exogenously increased from zero. This motivates the follow-up experiment, which directly tests these predictions by providing quality premiums to traders.

4.6.2 Stronger Quality-based Incentives at the Aggregation Stage

A higher α increases the value of quality for the MCC. As a result, MCCs may tighten acceptance standards, rejecting more low-quality milk and thereby increasing average quality $\bar{q}_m$ among accepted deliveries. Second, when pass-through is positive ($\lambda_i > 0$), higher α also raises farmer effort and individual quality q_i^*, further increasing $\bar{q}_m$ and any intermediary outcome that is increasing in realized quality or quality-linked revenue

$$\frac{\partial \bar{q_m}}{\partial \alpha} > 0 \tag{23}$$

4.6.3 Response to Threshold Changes under Quality-based Pricing

The model also generates predictions about how the behavioral response to quality-based pricing depends on the level of the acceptance threshold q_{min} relative to the prevailing quality distribution. When q_{min} is low relative to the natural distribution of quality (as when high-quality milk is abundant in the rainy season), most suppliers are comfortably above the threshold: $\pi_i \approx 1$ and $\partial \pi_i / \partial e_i \approx 0$. In this regime, the extensive margin contributes little to the marginal return to effort. The intensive margin $\pi_i \lambda_i \alpha \theta_i y_i$ is active and rewards both higher quality (through the price gradient) and higher volume (since accepted milk earns the quality premium on every liter). The model is agnostic about which of these responses dominates: this depends on the relative costs of sourcing additional volume versus sourcing higher-quality milk, which are not pinned down by the model.

When q_{min} is raised (as when quality thresholds are tightened or when the dry season reduces the natural quality distribution of available milk), a larger fraction of suppliers falls near or below the threshold. The acceptance probability π_i drops below one for many suppliers, and the extensive margin $\frac{\partial \pi_i}{\partial e_i}\left[p_0 y_i + \lambda_i \alpha \theta_i e_i y_i\right]$ strengthens. Here, the model predicts that raising the threshold increases the incentive to improve quality, since the threat of losing the entire bonus at the threshold creates a strong discrete incentive that complements the continuous gradient. Formally, the relative importance of the extensive margin in the first-order condition is increasing in q_{min}:

$$\frac{\partial}{\partial q_{min}} \left[\frac{\text{extensive margin}}{\text{total marginal return}}\right] > 0 \tag{24}$$

4.6.4 Farmer Responses to Stronger Downstream Incentives

With $\lambda_i > 0$, an increase in α raises farmers' optimal effort e_i^* and quality q_i^*, and also raises the quality-linked component of the price received.

$$\lambda_i > 0 \Rightarrow \frac{\partial e_i^*}{\partial \alpha} > 0 \, and \, \frac{\partial q_i^*}{\partial \alpha} > 0 \tag{25}$$

4.6.5 Summary of Testable Predictions

Table 6 summarizes the model's predictions, the channel and parameter variation each relies on, and the corresponding empirical test. The first two predictions concern the main experiment, where the introduction of measurement technology reduces σ^2 but, as the results show, α does not activate. The remaining predictions concern the follow-up experiment. We emphasize that the follow-up does not literally activate α as defined in the model, since α is the per-unit premium that the processor pays the MCC for verifiable quality; inducing $\alpha > 0$ would require an intervention at the processor stage, which the project could not implement. Instead, the follow-up injects a quality-contingent premium one node downstream, paying traders directly for the verified compositional quality of milk they deliver to the MCC. In the notation of the model this corresponds to setting the product $\lambda\alpha$ to a positive value at the trader-MCC interface by external transfer rather than letting it emerge from processor pricing. The resulting test is therefore not of whether $\alpha > 0$ propagates through the chain, but of whether a positive quality-linked transfer received by an intermediary alters its own quality behavior and is passed further upstream to farmers.

5 Follow-up Experiment: Activating the Price Channel

5.1 Design

The main experiment established that making quality observable through the measurement and digital monitoring system improved enforcement but did not, on its own, generate quality-based pricing. The follow-up experiment tests whether introducing explicit price incentives for quality targeted to traders can activate the price channel. The experiment builds directly on the milk testing and tracing infrastructure already deployed at MCCs as part of the main experiment.

The follow-up intervention targets traders, rather than MCCs or directly incentivizing farmers, for three reasons. First, qualitative data from the main experiment suggested that quality gains were smaller when milk passed through intermediaries than when farmers delivered directly to the MCC, indicating that the trader link is where the largest

Table 6: Model Predictions and Empirical Tests

Prediction	Channel	Parameter	Test	Result
		Main experiment: measurement technology ($\sigma^2 \to 0$)		
Adulteration falls	Enforcement	$\sigma^2 \to 0$	Milk samples	Fat +0.11pp, Water −0.49pp
Prices unchanged	Credibility, bargaining	$\sigma^2 \to 0$ insufficient for $\alpha > 0$	Farmer/MCC prices	All null
		Follow-up experiment: quality-based pricing ($\alpha > 0$)		
Quality and/or volume response	Intensive	$\alpha > 0$, low q_{min}	Stage 1	Qty +126L, Fat null
Stronger quality response	Extensive	$\alpha > 0$, high q_{min}	Stage 2	Fat +0.17pp
Limited pass-through	Price	$\alpha > 0$, $\lambda_i \approx 0$	Farmer prices	−17 UGX

unrealized quality improvements remain.[19] Second, a core objective of the follow-up is to study whether quality premiums are passed through to upstream suppliers. Providing the premium at the trader level creates a testable pass-through margin: we can observe whether traders share the bonus with the farmers from whom they source. Had we provided the premium directly to farmers, the price effect would be mechanical rather than informative about market transmission. Third, traders occupy a practical middle ground for statistical power. Providing premiums directly to MCCs would require randomizing across a prohibitively large number of collection centers to achieve adequate power, while incentivizing individual farmers would require paying premiums to roughly 500 subjects at considerably higher cost. Working with approximately 100 traders balances cost-effectiveness with sufficient statistical power to detect meaningful behavioral responses.

We focus on 19 MCCs in the Kazo district, a subset of the main experiment's treatment sites where the quality measurement and digital monitoring system remained operational. The geographic concentration within a single district ensures logistical feasibility to make sure the milk testing infrastructure remains operational throughout the study period; because identification relies on within-MCC comparisons, district-level characteristics are absorbed by the MCC fixed effects in the estimating equation (see Section 5.5 below). Within each MCC, traders who deliver milk were randomized to treatment or control, with randomization stratified by MCC, yielding 47 treated and 52 control traders. Both treated and control traders had their milk tested using the analyzer at each delivery, and both received quality feedback (fat and SNF readings) after each test. Treated traders additionally received a quality-contingent cash bonus via mobile money. This design isolates the price incentive from the information channel: any difference between groups reflects the effect of the financial reward for quality, not the effect of knowing one's quality readings.

The design exploits the high-frequency, repeated-transaction structure of dairy supply chains. Unlike staple crops where farmers often sell the bulk of their harvest in a single or only a few post-harvest transactions (Van Campenhout, Lecoutere, and D'Exelle, 2015), milk is collected and delivered daily, often in multiple transactions per day. This generates a dense panel of behavioral responses, allowing us to observe adjustment in near real time. A surprise mid-experiment threshold shift (from Stage 1 to Stage 2) provides within-subject variation in the stringency of quality requirements, allowing us to trace how the same traders adjust behavior as the quality bar tightens.

Prior to randomization, all traders were informed that the experiment involved a 50 percent chance of receiving quality-based bonuses and a 50 percent chance of serving as a comparison group. Informed consent was obtained under this framing. Because randomization is within MCC, control traders are aware that some of their peers receive bonuses. This raises the possibility of spillover effects, and the direction of bias differs across outcomes. For quality, if treated traders selectively source from the highest-quality farmers within the MCC catchment, control traders are left with a lower-quality pool. This sorting of existing quality across traders would inflate the treatment-control difference, biasing our quality estimates upward. For prices, the bias likely runs in the opposite direction: if treated traders compete more aggressively for farmers, farmgate prices rise for control-area farmers as well, attenuating the estimated price effect. The short duration of the experiment (31 days) limits the scope for large-scale supplier reallocation, but we cannot rule out these effects and note that our quality estimates should be interpreted as an upper bound on the true effect of the price incentive on quality improvement.

5.2 Intervention

Treated traders receive a quality premium of UGX 100 per liter for each 0.1 percentage point of fat and of UGX 100 per liter for each 0.1 percentage point of SNF content above the East African Community standards (which are 3.3 percent fat and 8.5 percent SNF). The digital infrastructure from the main experiment is essential to this intervention: it provides the verifiable, timestamped quality records on which bonus calculations are based.

The incentives are economically significant. The average control trader delivers approximately 270 liters per day with milk at 4.0 percent fat and 8.5 percent SNF. Under Stage 1 thresholds, such a trader would qualify for a daily bonus on the order of UGX 19,000 (roughly USD 5), a meaningful supplement to typical trading margins of about UGX 100 per liter. Among treated traders, the average daily bonus actually paid during Stage 1 was UGX 27,000 (median UGX 12,000).

---

[19] Focus group discussions suggest that farmers and traders respond to quality incentives along different margins. Farmers, who deliver a relatively fixed quantity determined by their herd size, tend to view milk quality as a matter of personal pride and have limited scope to adjust volume; their primary margin of response is quality improvement. Traders, by contrast, take a more commercial approach and can flexibly adjust both the volume they collect and the suppliers they source from, as long as deliveries clear the acceptance threshold. This asymmetry implies that quality premiums at the trader level may initially elicit volume rather than quality responses.

Bonuses are paid daily via mobile money, creating immediate and salient incentives that traders can link directly to each day's delivery quality. Each payment is accompanied by an SMS message reporting the trader's average fat and SNF content for that day's delivery along with the total bonus earned and volume delivered. Traders whose quality falls below the threshold instead receive an SMS informing them that their fat or SNF did not meet the minimum required to qualify for a bonus, along with the actual readings. This feedback loop ensures that traders observe not only whether they qualified but also how their compositional quality compares to the threshold, reinforcing the link between daily sourcing decisions and financial outcomes.

The experiment proceeds in two stages. In Stage 1 (which ran from November 1, 2025 through November 14, 2025), the quality thresholds are set at the East African Community standards: 3.3 percent fat and 8.5 percent SNF. Because the experiment coincided with the rainy season, when pasture quality is high and milk naturally exceeds these thresholds, nearly all treated traders qualified for the bonus. In terms of the model, Stage 1 represents a regime where q_{min} is low relative to the prevailing quality distribution: the extensive margin (acceptance probability) contributes little, and the intensive margin (the price gradient above the threshold) is the primary incentive. The model is agnostic about whether traders respond on quality, volume, or both.

In Stage 2 (beginning November 15, 2025), the fat threshold is raised without prior announcement to 3.9 percent while the SNF threshold remains unchanged. This increase shifts the regime: a substantial share of deliveries now falls near or below the new threshold, activating the extensive margin. The model predicts that quality-improving effort should strengthen in this regime, as the discrete threat of losing the entire bonus at the threshold creates a powerful incentive beyond the continuous gradient. The within-experiment threshold change allows us to observe how the same traders adjust behavior when conditions shift, providing a cleaner comparison of the two regimes than separate experiments would allow.

5.3 Sample and Timeline

Ex ante power calculations were conducted following the DeclareDesign framework (Blair et al., 2019). Results indicated that a sample of 100 traders with approximately 5 farmers per trader would provide adequate power to detect a 0.5 percentage point increase in fat content at the trader level, assuming an intra-cluster correlation of 0.11. More details can be found in a separate entry in the AEA RCT registry for the follow-up study.[20]

Stage 1 of the experiment began on November 1, 2025, with the threshold change marking Stage 2 taking effect on November 15. The experiment concluded on December 1, generating 5,596 daily milk quality submissions recorded through the digital monitoring system, comprising 2,622 trader-day observations: 1,224 in Stage 1 and 1,351 in Stage 2, with the remainder falling on boundary days. Endline surveys were conducted in the final days of the experiment, covering 99 of the 100 targeted traders and 422 farmers (197 in treated trader catchments and 225 in control).

5.4 Outcome Variables

We measure outcomes at three levels, drawing on both the digital monitoring system and purpose-built endline surveys.

The first source is the daily submission data recorded by the milk analyzer at each MCC, which yields the trader-day panel described above. For each delivery day we observe quantity-weighted average fat content (percent), quantity-weighted average solids-not-fat (SNF, percent), total liters delivered, and the quality bonus earned (UGX). These readings are produced by the milk analyzer and submitted by MCC managers through the digital monitoring system. While this avoids the recall bias inherent in survey-based quality measures, the data depend on the manager correctly testing each delivery and entering the results.

The second source is the trader endline survey. Key outcomes include liters delivered on the day preceding the interview, whether any milk was rejected in the preceding month, whether the trader pays farmers a premium for quality, and the volume-weighted average price paid to supplying farmers (UGX per liter).

The third source is the farmer endline survey. Outcomes include the volume-weighted price received from the trader (UGX per liter, constructed from up to six recent transactions), whether quality was checked in any recent transaction, and a set of feeding practice indicators: use of bran, use of crop residues, use of mineral licks, and use of controlled grazing, which are combined into an Anderson index to summarize investments in milk quality at the farm level.

---

[20] The pre-analysis plan was pre-registered at the AEA RCT registry (https://www.socialscienceregistry.org/trials/16750)

5.5 Estimation and Inference

For the daily submission data, we estimate treatment effects using OLS with MCC fixed effects:

$$y_{jt} = \alpha + \beta \, \text{Treatment}_j + \mu_m + \varepsilon_{jt} \tag{26}$$

where y_{jt} is the outcome for trader j on day t, Treatment_j is an indicator equal to one if trader j was assigned to receive quality bonuses, and μ_m is a fixed effect for the MCC at which trader j delivers. Standard errors are clustered at the trader level. We estimate this specification over three windows: the full post-treatment period (November 3 onward), Stage 1 only (November 3 through 14), and Stage 2 only (November 15 onward). Comparing coefficients across stages reveals whether traders adjust to the higher quality threshold.

For the trader endline, we estimate an analogous specification with MCC fixed effects but without clustering, since traders are themselves the unit of randomization. For the farmer endline, the specification again includes MCC fixed effects, with standard errors clustered at the trader level to account for the nesting of farmers within trader catchments.

As was the case in the main experiment, our primary inferential framework is Fisher Randomization Inference, computed from 5,000 permutations of the treatment indicator within MCC blocks. To match the fixed-effects specification, outcomes are demeaned at the MCC level before computing the RI distribution. To account for multiple testing across outcomes within a family, we report Westfall-Young stepdown-adjusted p-values, which control the familywise error rate while exploiting the correlation structure across outcomes.

5.6 Attrition and Balance

Attrition is minimal. The daily submission data are administrative records from the digital monitoring system and do not suffer from survey nonresponse. For the endline surveys, 99 of the 100 targeted traders were interviewed, and 422 of the approximately 500 targeted farmers completed the questionnaire (197 in treated trader catchments and 225 in control). The shortfall in the farmer sample reflects the difficulty of locating smallholder farmers who sell through mobile traders rather than differential attrition by treatment status.

To verify that randomization produced comparable samples, Appendix Table A10 reports balance for the 99 traders on pre-treatment quality and quantity outcomes and on plausibly pre-determined trader characteristics, and Appendix Table A11 reports balance for the 422 farmers on plausibly pre-determined production and trader-linkage characteristics. Pre-treatment fat content is approximately 0.24 percentage points lower among traders subsequently assigned to treatment (significant at the 1 percent level), while pre-treatment SNF and quantity are balanced. Because our daily-submission specification includes MCC fixed effects but does not net out trader-level pre-treatment outcomes, we estimate a robustness specification that adds the trader-specific pre-treatment mean fat as a control; the estimated treatment effect on post-treatment fat shifts only marginally, from 0.109 percentage points in the main specification to 0.096 percentage points (significant at the 5 percent level) when conditioning on pre-treatment levels, indicating that the result is not driven by mean reversion from the lower pre-treatment baseline. One trader characteristic also differs at conventional levels: traders in the treatment arm are somewhat more likely to be owner-operators rather than employees of a larger trading firm, a pre-determined feature that the MCC fixed effects in our main specification partially absorb.

5.7 Results

5.7.1 Daily Submission Outcomes

Table 7 reports treatment effects on daily milk submissions for the full experimental period (Panel A) and separately for Stage 1 (Panel B, November 3 to 14, fat threshold of 3.3 percent) and Stage 2 (Panel C, November 15 onward, fat threshold of 3.9 percent). All specifications include MCC fixed effects with standard errors clustered at the trader level. We report randomization inference p-values and Westfall-Young adjusted p-values to account for multiple testing across the four outcomes.

Over the full period, treated traders deliver milk with 0.109 percentage points higher fat content (RI $p = 0.003$, WY $p = 0.007$) and 0.030 percentage points higher solids-not-fat (RI $p = 0.119$). Treated traders also deliver 124 liters more per day (RI $p = 0.013$, WY $p = 0.025$). To summarize the economic magnitude of these behavioral responses, we compute the bonus for all traders (including a "shadow" bonus for the control) by applying the bonus formula to each day's quality readings and volume. The treatment effect on this bonus is 13,000 UGX per day (RI

Table 7: Follow-up Experiment: Treatment Effects on Daily Milk Quality

	Fat (%)	SNF (%)	Quantity (L)	Bonus (1000 UGX)
	Panel A: Full period			
treat	0.108^{**}	0.029	124.3^{*}	13^{**}
	(0.031)	(0.016)	(48.51)	(3.9)
N	2425	2425	2425	2425
ctrl	3.95	8.53	273.7	14.3
	Panel B: Stage 1 (fat threshold 3.3%)			
treat	0.035	0.004	126.1^{**}	13.8^{**}
	(0.027)	(0.023)	(39.18)	(4.3)
N	1074	1074	1074	1074
ctrl	3.92	8.54	268.7	20.7
	Panel C: Stage 2 (fat threshold 3.9%)			
treat	0.169^{**}	0.051^{*}	122.9^{*}	12.7^{**}
	(0.041)	(0.016)	(59.04)	(4)
N	1351	1351	1351	1351
ctrl	3.98	8.53	277.6	9.3

Note: Each panel reports OLS estimates with MCC fixed effects and trader-clustered standard errors. The Bonus column reports a shadow bonus computed by applying the quality premium formula to each trader's daily readings and volume, including for control traders who did not actually receive payments. The treatment effect on this variable captures the additional bonus that treated traders would earn due to their behavioral response. RI = randomization inference p-value; WY = Westfall-Young stepdown-adjusted p-value. Significance: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ (RI p-values).

p = 0.001, WY p = 0.004), the additional daily earnings that treated traders' quality and volume improvements would generate relative to control traders facing the same pricing schedule. The fat and quantity effects are robust to multiple testing correction; the SNF effect is not.

The contrast between Stage 1 and Stage 2 reveals the central finding of the follow-up experiment. During Stage 1, when the fat threshold is set at 3.3 percent (a level easily met during the rainy season, when feed is abundant and baseline fat content is high), treated traders show no significant improvement in quality: the fat effect is 0.035 percentage points (RI p = 0.263) and the SNF effect is 0.004 percentage points (RI p = 0.888). Yet even in Stage 1, treated traders deliver significantly more milk (126 liters, almost 50 percent above the control mean, RI p = 0.004, WY p = 0.010) and earn significantly higher bonuses (13,784 UGX, RI p = 0.003, WY p = 0.009). This pattern is consistent with traders responding to the extensive margin of the bonus: when the quality threshold is easy to clear, the optimal response is to source more milk (from additional farmers or larger quantities from existing suppliers) rather than to screen for quality.

When the threshold increases to 3.9 percent at the start of Stage 2, the behavioral response shifts markedly. The fat content effect jumps to 0.169 percentage points (RI $p < 0.001$, WY $p < 0.001$) and SNF rises by 0.051 percentage points (RI p = 0.014, WY p = 0.028), both highly significant even after multiple testing adjustment. The quantity and bonus effects remain significant at 123 liters (RI p = 0.042, WY p = 0.042) and 12,743 UGX (RI p = 0.001, WY p = 0.004), respectively. The emergence of strong quality effects only after the threshold increase maps directly onto the model's prediction about the extensive margin: as the minimum quality standard $q_{\min}$ rises, the set of farmers whose milk clears the threshold shrinks, and the returns to selective sourcing increase. Traders shift from a volume strategy (source broadly, rely on naturally high quality) to a quality strategy (screen suppliers, favor those with higher-fat milk).

The daily panel reveals adjustment dynamics that would be invisible in a standard endline-only design. Figures 5 and 6 plot average quantity and daily fat content by treatment status over the 30-day experiment. The volume response in Stage 1 is immediate: within the first few days of the bonus activation, treated traders deliver visibly more milk than control traders, and this gap is sustained throughout the period. The absence of a corresponding quality response in Stage 1 is consistent with volume expansion being a lower-cost margin of adjustment than sourcing higher-quality milk: traders can recruit additional suppliers or collect larger quantities from existing ones

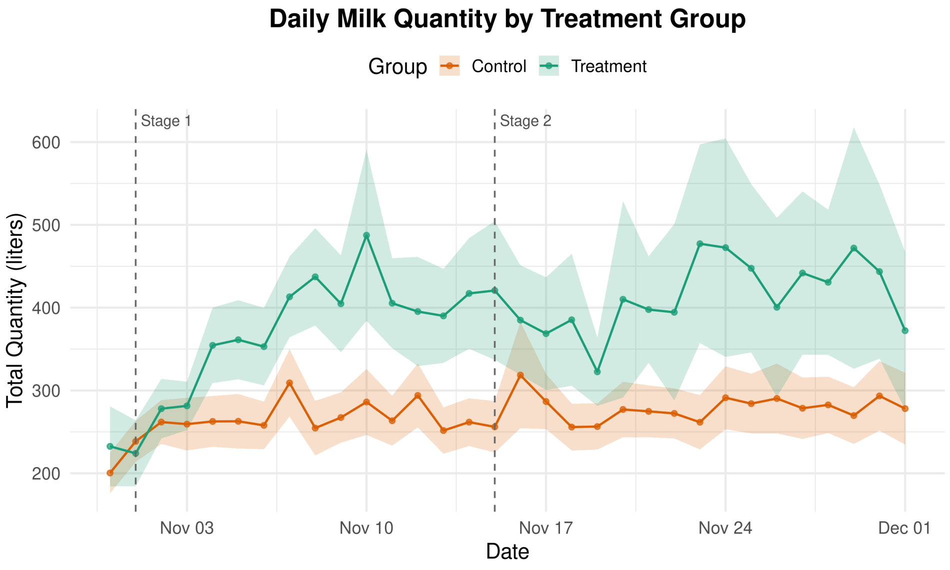


Figure 5: Daily Average Quantity by Treatment Status

without fundamentally changing their sourcing network. After the threshold change on November 15, the quality response emerges more gradually, with the fat content gap between treatment and control widening visibly only after about a week. This slower adjustment is consistent with quality sourcing requiring more costly restructuring of supplier networks, such as identifying which farmers produce higher-fat milk and redirecting collection accordingly.

5.7.2 Trader Endline Survey

Despite the strong effects in the daily submission data, the trader endline survey conducted during the final days of the experiment shows no significant treatment effects across any of the four outcomes: total liters delivered to the MCC on the day preceding the interview (RI p = 0.396), number of times the trader rejected farmer milk due to poor quality in the preceding month (RI p = 0.691), whether the trader paid a higher price to farmers supplying better-quality milk in the preceding week (RI p = 0.155), and the volume-weighted average price per liter paid to farmers, constructed from up to six individual farmer-level purchase records (RI p = 0.413). None of the four outcomes approaches conventional significance levels (Table 8).

Two factors likely explain the divergence between the daily and survey results. First, the trader sample is small (99 traders), limiting statistical power for survey-based detection. Second, the daily submission data aggregate behavior over 31 days of repeated decisions, whereas the endline captures a single cross-sectional snapshot. Cumulative treatment effects visible in panel data can fall below detection thresholds in a one-time survey. Notably, the endline was conducted during the final days of the experiment while the bonus system was still operational, so the null survey results cannot be attributed to the treatment having ended.

5.7.3 Farmer-Level Outcomes

We also examine whether the trader-level treatment generates upstream effects on the farmers from whom treated traders source milk (Table 9). The most striking finding is a negative effect on farmer prices: farmers selling to treated traders report receiving 16.89 UGX less per liter (RI p = 0.017, WY p = 0.050). This negative pass-through is significant and runs counter to the prediction that quality premiums at the trader level would raise prices for farmers.

Other farmer outcomes provide limited evidence of behavioral transmission. Whether the farmer reports having quality checked is unaffected by treatment (coefficient of -0.027, not significant). The feeding practices index shows a marginally significant improvement of 0.134 (RI p = 0.076), though this does not survive Westfall-Young multiple testing correction (WY p = 0.145). No individual feeding practice is significant on its own. The marginal feeding

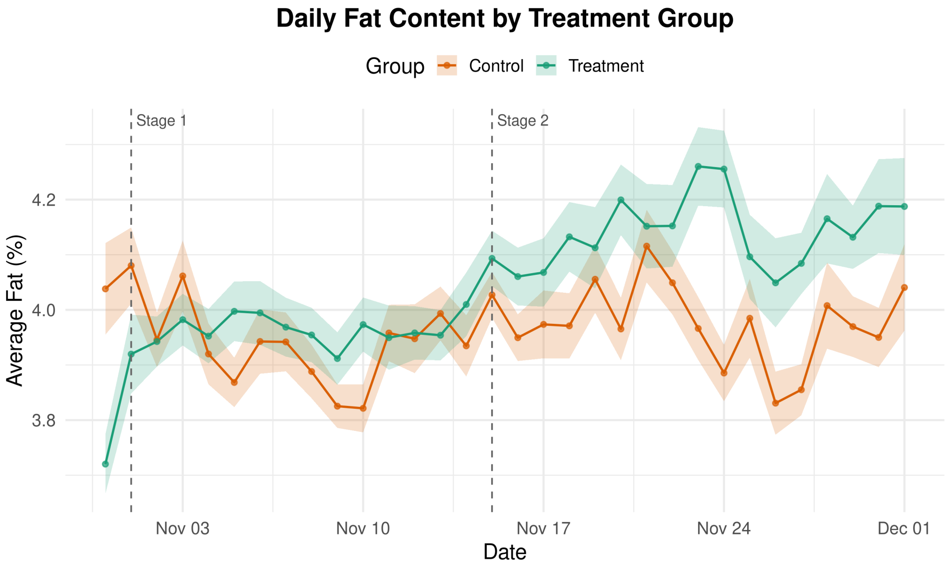


Figure 6: Daily Average Fat Content by Treatment Status

Table 8: Follow-up Experiment: Trader Endline Survey Results

	Control Mean	Treatment Effect	N
Delivered quantity (liters)	389.3	-36.26	99
		(40.62)	
Any milk rejected (past month)	0.500	-0.039	99
		(0.079)	
Pays quality premium to farmers	0.058	0.088	99
		(0.061)	
Average purchase price (UGX/liter)	899.4	-6.80	99
		(7.84)	
Anderson index	0.003	-0.007	99
		(0.086)	

Notes: OLS estimates with CR2 cluster-robust standard errors in parentheses. Significance stars from randomization inference p-values: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 9: Follow-up Experiment: Farmer level outcomes

	Control Mean	Treatment Effect (SE)	N
Average price received (UGX/liter)	906.1	-16.89* (6.36)	422
Quality checked by trader	0.618	-0.027 (0.061)	422
Feeding practices index	-0.069	0.134^{+} (0.064)	422
Used bran/concentrates	0.093	-0.002 (0.031)	422
Used crop residues	0.093	0.054 (0.032)	422
Used salt/mineral lick	0.511	0.081 (0.061)	422
Controlled grazing	0.364	0.093 (0.054)	422
Anderson index	0.016	-0.036 (0.059)	422

Notes: OLS estimates with CR2 cluster-robust standard errors in parentheses. Significance stars from randomization inference p-values: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

index result suggests that some information about quality requirements may transmit from treated traders to their farmers, inducing modest adjustments in feeding behavior even in the absence of price incentives.

6 Discussion and Conclusion

This paper asked whether making quality observable and traceable can give rise to a market for quality in smallholder value chain where quality discovery is costly and aggregation happens early in the chain. We studied the question in two linked field experiments in the context of Uganda's dairy sector. The first, a cluster-randomized trial across 95 catchment areas, introduced milk analyzers and digital record-keeping at collection centers; it improved milk quality, primarily by curbing adulteration, but produced no price premiums anywhere along the chain, because processors purchase at uniform per-liter rates and transmit no demand for quality downstream. The second experiment injected the missing price signal directly, randomly paying quality premiums to traders; this shifted trader behavior toward quality-based sourcing once the standard became binding, yet the resulting surplus was retained within the intermediation layer rather than passed upstream to farmers. Together the two experiments show that observability is necessary but not sufficient for a market for quality to emerge: even when prices are made to reward quality, the structure of intermediation governs who captures the returns.

Three of these findings merit closer interpretation: the dominance of the threshold over the gradient in the follow-up experiment, the negative effect on farm-gate prices, and the broader relationship between the two experiments.

Traders in the follow-up experiment responded to the threshold, not the gradient. When a low threshold made the quality bonus easy to earn (Stage 1), traders expanded volume without improving quality. When the threshold rose and most deliveries failed to qualify (Stage 2), traders shifted to quality-based sourcing, with fat content increasing significantly. The continuous price gradient, which rewarded marginal quality improvements throughout both stages, elicited no detectable response. This asymmetry suggests that discrete pass/fail signals are more effective than smooth incentives at triggering costly behavioral change among intermediaries.

Three mechanisms could generate this pattern. First, loss aversion: traders who earned the bonus in Stage 1 may perceive the higher threshold as a loss, motivating costly adjustment to restore the payment. The approximately one-week delay in quality response (Figure 6) does not rule out this channel, because even loss-averse traders need time to identify new suppliers and restructure collection routes. Second, salience: the binary qualify/fail signal is cognitively simpler than computing marginal returns along a continuous schedule, and may command more attention. Third, cost-based adjustment: restructuring supplier networks involves lumpy fixed costs, and the gradient's smooth marginal returns may fall below the action threshold at every point. The gradual emergence of quality improvements is consistent with all three mechanisms, and the evidence does not allow us to distinguish among them conclusively.

These findings have implications for incentive design, though we note that they rest on one experiment with a single threshold change, not a direct comparison of threshold and gradient structures. That said, the pattern suggests that when intermediaries face lumpy adjustment costs, threshold-based designs may outperform continuous gradients by concentrating stakes at a single point where the return to reorganization exceeds the fixed cost. Programs seeking to activate quality upgrading through supply chain intermediaries may benefit from incorporating discrete performance targets rather than relying solely on linear price schedules.

The negative effect on farm-gate prices is another striking finding and warrants careful interpretation. Three mechanisms may contribute. First, a composition effect: treated traders who shift toward quality-based sourcing may reallocate purchases toward higher-volume farmers who accept lower per-unit prices, changing the composition of sampled transactions. Second, bargaining leverage: traders with quality information from the milk analyzer gain an informational advantage over farmers, potentially extracting surplus in price negotiations. Third, selection on outside options: higher-quality milk may be concentrated in remote, less market-integrated areas where farmers face fewer competing buyers and therefore command weaker prices. Treated traders screening for quality would then source disproportionately from these low-competition pockets, lowering the average price observed among their suppliers even without any change in the price offered to a given farmer. These channels are not mutually exclusive, and the short duration of the follow-up experiment limits our ability to assess their relative importance. The finding underscores that quality-contingent incentives introduced at one node of the chain can redistribute surplus in ways that disadvantage upstream actors, and deserves further investigation in longer-run settings.

Mapped onto the model, the two experiments trace its central parameter, the share of quality that is priced: the main experiment holds it at zero, and the follow-up raises it above zero by fiat. The contrast isolates the paper's theoretical point. Information resolves the asymmetry at the node where the technology is deployed, but unless quality measurement is tied to payment at every node of the chain, putting a price on quality merely relocates the surplus to whichever intermediary holds the information rather than delivering it to the producer whose effort created the quality.

These findings connect to, and sharpen, several active literatures. On the informational preconditions for quality markets, our result that large gains in measurement raise quality through enforcement while leaving prices untouched echoes Rao and Shenoy (2023), where aggregate quality incentives in Indian dairy cooperatives improve cleanliness without quality-based farmer pay, and Xu (2024), where digital traceability in Kenyan dairy curbs adulteration through monitoring rather than new price signals. By randomizing observability and pricing separately, we isolate what these studies bundle: observability is necessary but not sufficient, and the absence of downstream demand, as in the near-zero return to quality that Bold et al. (2022) document for Ugandan maize, is the binding constraint. On pass-through, our finding that an exogenous trader premium is captured rather than transmitted aligns with Bergquist and Dinerstein (2020), who show that intermediaries retain the bulk of surplus, and with Macchiavello et al. (2026) and Bai et al. (2025), where quality premiums attenuate upstream under intermediary buyer power. It contrasts with Casaburi and Reed (2022), where an unconditional trader subsidy in Sierra Leone cocoa passes through to farmers under competition: the divergence points to the information structure of the incentive, since a private, quality-contingent bonus paid on readings that farmers cannot observe creates no competitive pressure to share the rent.

For policy, measurement infrastructure has clear standalone value: it enforces a quality floor, and the 24 percent adoption rate among control MCCs by endline reveals genuine demand for it as a compliance tool. But a floor is not a market. Rewarding quality at the top requires coordinated action on margins a single technology cannot reach: downstream demand for quality and contractual infrastructure linking individual performance to pay at each node. In Uganda's dairy chain that gap is concrete, at the processor-MCC interface, where uniform per-liter pricing leaves MCCs no quality signal to pass to farmers. Until it is closed, measurement will enforce quality but not price it.

References

Akerlof, G. A. 1970. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84 (3): 488–500.

Anderson, M. L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–1495.

Antràs, P. and D. Chor. 2013. "Organizing the Global Value Chain." *Econometrica* 81 (6): 2127–2204.

Ashraf, N., X. Giné, and D. Karlan. 2009. "Finding Missing Markets (and a Disturbing Epilogue): Evidence from an Export Crop Adoption and Marketing Intervention in Kenya." *American Journal of Agricultural Economics* 91 (4): 973–990.

Atkin, D., A. K. Khandelwal, and A. Osman. 2017. "Exporting and Firm Performance: Evidence from a Randomized Experiment." *Quarterly Journal of Economics* 132 (2): 551–615.

Bai, J. 2025. "Melons as Lemons: Asymmetric Information, Consumer Learning and Seller Reputation." *The Review of Economic Studies* 92 (6): 3574–3610.

Bai, J., L. F. Bergquist, A. Morjaria, R. Morton, and Y. Tang. 2025. "Quality Incentives and Upgrading in Uganda's Coffee Supply Chain." Working Paper.

Bergquist, L. F. and M. Dinerstein. 2020. "Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya." *American Economic Review* 110 (12): 3705–3747.

Blair, G., J. Cooper, A. Coppock, and M. Humphreys. 2019. "Declaring and diagnosing research designs." *American Political Science Review* 113 (3): 838–859.

Bold, T., S. Ghisolfi, F. Nsonzi, and J. Svensson. 2022. "Market Access and Quality Upgrading: Evidence from Four Field Experiments." *American Economic Review* 112 (8): 2518–2552.

Casaburi, L. and T. Reed. 2022. "Using Individual-Level Randomized Treatment to Learn about Market Structure." *American Economic Journal: Applied Economics* 14 (4): 58–90.

Fieler, A. C., M. Eslava, and D. Y. Xu. 2018. "Trade, Quality Upgrading, and Input Linkages: Theory and Evidence from Colombia." *American Economic Review* 108 (1): 109–146.

Hansman, C., J. Hjort, G. León-Ciliotta, and M. Teachout. 2020. "Vertical Integration, Supplier Behavior, and Quality Upgrading among Exporters." *Journal of Political Economy* 128 (9): 3358–3413.

Hoffmann, V., S. Kariuki, J. Pieters, and M. Treurniet. 2023. "Upside Risk, Consumption Value, and Market Returns to Food Safety." *American Journal of Agricultural Economics* 105 (3): 914–939.

Holmström, B. 1982. "Moral Hazard in Teams." *Bell Journal of Economics* 13 (2): 324–340.

Imbens, G. W. and M. Kolesár. 2016. "Robust Standard Errors in Small Samples: Some Practical Advice." *Review of Economics and Statistics* 98 (4): 701–712.

Kugler, M. and E. Verhoogen. 2012. "Prices, Plant Size, and Product Quality." *The Review of Economic Studies* 79 (1): 307–339.

Macchiavello, R., J. Miquel-Florensa, N. de Roux, E. Verhoogen, M. Bernasconi, and P. W. Farrell. 2026. "Quality Upgrading in Global Supply Chains: Evidence from Colombian Coffee." NBER Working Paper No. 34610.

Minten, B. and T. Reardon. 2008. "Food Prices, Quality, and Quality's Pricing in Supermarkets versus Traditional Markets in Developing Countries." *Review of Agricultural Economics* 30 (3): 480–490.

Minten, B., L. Randrianarison, and J. F. M. Swinnen. 2009. "Global Retail Chains and Poor Farmers: Evidence from Madagascar." *World Development* 37 (11): 1728–1741.

Park, S., Z. Yuan, and H. Zhang. 2025. “Technology Training, Buyer-Supplier Relationship, and Quality Upgrading in an Agricultural Supply Chain.” *The Review of Economics and Statistics* 107 (3): 711–727.

Rao, M. and A. Shenoy. 2023. “Got (Clean) Milk? Organization, Incentives, and Management in Indian Dairy Cooperatives.” *Journal of Economic Behavior & Organization* 212: 708–722.

Saenger, C., M. Qaim, M. Torero, and A. Viceisza. 2013. “Contract Farming and Smallholder Incentives to Produce High Quality: Experimental Evidence from the Vietnamese Dairy Sector.” *Agricultural Economics* 44 (3): 297–308.

Shapiro, C. 1983. “Premiums for High Quality Products as Returns to Reputations.” *Quarterly Journal of Economics* 98 (4): 659–679.

Tirole, J. 1996. “A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality).” *Review of Economic Studies* 63 (1): 1–22.

Treurniet, M. 2021. “The Potency of Quality Incentives: Evidence from the Indonesian Dairy Value Chain.” *American Journal of Agricultural Economics* 103 (5): 1661–1678.

Van Campenhout, B., E. Lecoutere, and B. D’Exelle. 2015. “Inter-temporal and spatial price dispersion patterns and the well-being of maize producers in Southern Tanzania.” *Journal of African Economies* 24 (2): 230–253.

Van Campenhout, B., B. Minten, and J. F. M. Swinnen. 2021. “Leading the Way: Foreign Direct Investment and Dairy Value Chain Upgrading in Uganda.” *Agricultural Economics* 52 (4): 607–631.

Verhoogen, E. 2023. “Firm-Level Upgrading in Developing Countries.” *Journal of Economic Literature* 61 (4): 1410–1464.

Wanjohi, M., L. Gitonga, K. Marshall, G. Njoroge, A. I. Abdi, and B. Bett. 2020. “Milk Quality along Dairy Farming Systems and Associated Value Chains in Kenya: An Analysis of Composition, Contamination and Adulteration.” *Food Control* 119: 107482.

Westfall, P. H. and S. S. Young. 1993. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. New York: Wiley.

Winfree, J. A. and J. J. McCluskey. 2005. “Collective Reputation and Quality.” *American Journal of Agricultural Economics* 87 (1): 206–213.

Xu, G. 2024. “Farm to Fridge: Digital Traceability and Quality Upgrading in the Kenyan Dairy Value Chain.” Available at SSRN 5722923.

Online Appendix

Table A1: Additional Balance of MCC Baseline Characteristics

	ctrl	treat	nobs
Full-time employees (number)	2.97	0.346	124
	(1.89)	(0.337)	
Number of farmers/traders supplying milk (rainy season)	55.92	-6.57	112
	(64.04)	(10.03)	
Capacity utilization in dry season (%)	37.98	-4.56	119
	(20.99)	(3.72)	
Number of milk cans owned	21.05	0.716	124
	(52.96)	(7.88)	
Provides credit to suppliers (1=yes)	0.833	0.057	124
	(0.376)	(0.070)	
F-statistic		1.60	
p-value		0.269	

Note: This table reports additional baseline balance between MCCs assigned to the measurement-system treatment and those in the control group. Column 1 presents control group means with standard deviations in parentheses. Column 2 shows the difference in means (treatment minus control) with robust standard errors in parentheses. Column 3 lists the number of non-missing observations. The F-statistic and p-value correspond to a joint test of equality of all covariates. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table A2: Additional Balance of Farmer Baseline Characteristics

	ctrl	treat	nobs
Household size (number of members)	9.75	-0.512	2261
	(4.69)	(0.342)	
Liters milk produced per day (rainy season)	70.05	9.44	2261
	(61.48)	(7.61)	
Sells to MCC in rainy season (1=yes)	0.793	-0.023	2261
	(0.405)	(0.049)	
Uses steel containers for delivery (1=yes)	0.752	0.046	2261
	(0.432)	(0.039)	
Member of dairy cooperative (1=yes)	0.757	-0.084^{+}	2261
	(0.429)	(0.048)	
F-statistic		1.28	
p-value		0.288	

Note: This table reports additional baseline balance between farmers connected to treatment MCCs and those connected to control MCCs. Column 1 presents control group means with standard deviations in parentheses. Column 2 shows the treatment-control difference with CR2 cluster-robust standard errors in parentheses, clustered at the catchment area level. Column 3 gives the number of non-missing observations for each variable. The F-statistic and p-value correspond to a joint test of equality of all covariates. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table A3: Milk quality (LATE estimates)

	ctrl	treat (LATE)	nobs
Butter fat (%)	3.88	0.395^{+}	2491
	(0.539)	(0.204)	
SNF (%)	8.58	0.247	2491
	(0.490)	(0.171)	
Added Water (%)	1.63	-1.53	2491
	(3.68)	(1.04)	
Protein (%)	3.16	0.091	2491
	(0.194)	(0.072)	
Density (CLR)	27.93	1.29^{+}	2491
	(2.31)	(0.737)	
Index	-0.103	0.675*	2491
	(0.817)	(0.319)	

Note: Column 1 reports mean milk quality measures for the control group, with standard deviations in parentheses. Column 2 presents second-stage 2SLS estimates instrumenting actual analyzer use with randomized treatment assignment, with robust standard errors in parentheses. The first stage is not separately reported; the first-stage F-statistic for instrument relevance is 262.7, well above conventional weak-instrument thresholds. Column 3 shows the number of non-missing observations. The index is an Anderson (2008) index of the five quality components, with added water entering with a negative sign. Standard errors are clustered at the catchment-area level. This table reports LATE estimates corresponding to the ITT results in Table3. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table A4: Primary farmer outcomes (LATE estimates)

	ctrl	treat (LATE)	nobs
Improved Practices Index	0.014	-0.056	2054
	(0.562)	(0.113)	
Buyer Checks with MA	0.157	0.197	1337
	(0.364)	(0.131)	
Average Sales Price	1020.1	32.51	1254
	(97.59)	(29.19)	
Gets Quality Bonus	0.081	-0.042	1281
	(0.274)	(0.062)	
Index	-0.036	0.220^{+}	1202
	(0.490)	(0.132)	

Note: Column 1 reports control group means with standard deviations in parentheses. Column 2 presents second-stage 2SLS estimates instrumenting actual analyzer use with randomized treatment assignment, with robust standard errors in parentheses. The first stage is not separately reported; first-stage F-statistics for instrument relevance range from 178 to 300 across outcomes, well above conventional weak-instrument thresholds. Column 3 shows the number of observations. The index is an Anderson (2008) index of the four primary outcomes. Standard errors are clustered at the catchment-area level. This table reports LATE estimates corresponding to the ITT results in Table4. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table A5: Primary MCC outcomes (LATE estimates)

	ctrl	treat (LATE)	nobs
Tests Incoming (MA)	0.288	1.44**	120
	(0.457)	(0.245)	
Tests Outgoing (MA)	0.203	0.781**	120
	(0.406)	(0.230)	
Price Bought	1075.0	-42.06	115
	(92.56)	(47.09)	
Avg Sales Price	1199.6	9.89	108
	(106.3)	(55.39)	
Gives Quality Bonus	0.186	-0.086	119
	(0.393)	(0.191)	
Gets Quality Bonus	0.186	0.084	119
	(0.393)	(0.195)	
Index	-0.077	0.467*	115
	(0.477)	(0.233)	

Note: Column 1 reports control group means with standard deviations in parentheses. Column 2 presents second-stage 2SLS estimates instrumenting actual analyzer use with randomized treatment assignment, with robust standard errors in parentheses. The first stage is not separately reported; first-stage F-statistics for instrument relevance range from 14.9 to 19.1 across outcomes, above conventional weak-instrument thresholds. Column 3 shows the number of observations. The index is an Anderson (2008) index of the six primary MCC outcomes. Standard errors are clustered at the catchment-area level. This table reports LATE estimates corresponding to the ITT results in Table5. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table A6: Secondary outcomes at farmer level - uptake

	ctrl	treat	nobs
Remembers video	0.209	-0.006	2059
	(0.407)	(0.031)	
Remembers receiving seed	0.251	0.003	2059
	(0.434)	(0.029)	
Used seed	0.171	-0.005	2059
	(0.377)	(0.028)	
Knows compositional quality matters	-0.037	0.038	2059
	(0.767)	(0.069)	
Index of uptake	-0.211	0.015	2059
	(0.571)	(0.041)	

Note: Column 1 reports control group means with standard deviations in parentheses. Column 2 presents the treatment effect of the MCC-level measurement-system intervention, with robust standard errors in parentheses. Column 3 shows the number of observations. The index is an Anderson (2008) index of the four uptake outcomes. Standard errors are clustered at the catchment-area level. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table A7: Uptake and Operational Use of Milk Analyzers at MCCs

	ctrl	treat	nobs
Poster is visible (1=yes)	0.034	0.343^{**}	120
	(0.183)	(0.076)	
Milk Analyzer present (1=yes)	0.237	0.533^{**}	120
	(0.429)	(0.081)	
Project Milk Analyzer is present (1=yes)	0.085	0.637^{**}	120
	(0.281)	(0.076)	
Milk analyzer operational (1=yes)	0.237	0.369^{**}	120
	(0.429)	(0.086)	
MCC uses digital record-keeping (1=yes)	0.237	0.369^{**}	120
	(0.429)	(0.087)	
Index of uptake	-0.436	0.858^{**}	120
	(0.440)	(0.118)	

Note: olumn 1 reports the mean of each outcome for the control group, with standard deviations in parentheses. Column 2 shows the estimated treatment effect of the measurement-system intervention at the MCC level, with robust standard errors in parentheses. Column 3 lists the number of MCCs with non-missing observations. All prices are in Ugandan shillings (UGX). The "Index of uptake" is an Anderson (2008) index of the variables listed above. Significance levels: + p < 0.10, * p < 0.05, ** p < 0.01.

Table A8: Attrition correlates, Experiment 1 farmer sample

	retained	attrited	diff.	SE	N
Household size	9.58	9.49	-0.089	(0.405)	2261
Household head age	54.38	51.70	-2.68^{+}	(1.38)	2261
Herd size	65.15	80.53	15.38^{+}	(7.76)	1976
Share improved breed	0.931	0.914	-0.017	(0.015)	1976
Liters/day (wet)	75.52	81.64	6.12	(7.26)	2261
Liters sold/day (wet)	63.77	70.10	6.33	(5.96)	2261
Sells to MCC (wet)	0.795	0.754	-0.041	(0.045)	2261
Uses steel container	0.777	0.788	0.011	(0.049)	2261
Cooperative member	0.721	0.635	-0.086	(0.053)	2261
Acaricide expenditure	71.79	72.72	0.927	(9.81)	904
Joint F on all covariates (HTZ)	$F = 1.55$, $p = 0.160$				789

Note: Comparison of baseline characteristics between farmers retained in the endline analysis sample (N=2,058) and those lost to attrition (N=203, 9.0 percent of the baseline sample). Columns 1-2 report unconditional means; column 3 reports the difference (attrited minus retained) with cluster-robust standard errors at the catchment level (CR2). The joint test is a Hotelling-type Wald test (HTZ) of the null that all 10 baseline characteristics jointly fail to predict attrition; listwise deletion is used and N reflects observations with complete data on all covariates (Acaricide expenditure has substantial missingness in the baseline). Significance levels: + p < 0.10, * p < 0.05, ** p < 0.01.

Table A9: Balance on Experiment 1 farmer analysis sample

	ctrl	treat	N
Household size	9.64	-0.114	2058
	(4.57)	(0.317)	
Household head age	54.67	-0.569	2058
	(13.37)	(0.777)	
Herd size	64.98	0.314	1800
	(61.84)	(5.73)	
Share improved breed	0.936	-0.010	1800
	(0.170)	(0.012)	
Liters/day (wet)	74.16	2.65	2058
	(65.12)	(7.13)	
Liters sold/day (wet)	63.45	0.620	2058
	(59.70)	(5.96)	
Sells to MCC (wet)	0.805	-0.020	2058
	(0.396)	(0.046)	
Uses steel container	0.764	0.026	2058
	(0.425)	(0.037)	
Cooperative member	0.770	-0.097*	2058
	(0.421)	(0.047)	
Acaricide expenditure	73.57	-3.50	830
	(122.9)	(10.04)	
Joint F on all covariates (HTZ)		$F = 0.819$, $p = 0.612$	725

Note: Re-estimation of baseline balance on the Experiment 1 analysis sample (farmers retained at endline, N=2,058) rather than on the full baseline sample. Each row reports the pure control (T1=0 T2=0) mean and standard deviation, the pooled measurement-system coefficient (orthogonalised with respect to the second randomised arm) with CR2 cluster-robust standard errors at the catchment level, and the observation count. The joint test is a Hotelling-type Wald test (HTZ) of the null that all baseline characteristics jointly fail to predict treatment assignment in the analysis sample; listwise deletion is used and N reflects observations with complete data on all covariates. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table A10: Experiment 2 trader balance

	ctrl	treat	SE	N
Pre-treatment fat (%)	4.06	-0.239**	(0.064)	114
Pre-treatment SNF (%)	8.55	-0.076	(0.048)	114
Pre-treatment quantity (litres/day)	223.0	-1.16	(31.06)	114
Pre-treatment bonus (UGX/day)	19725.6	-5535.9^{+}	(3045.0)	114
Trader characteristics				
Self (owner-trader, $\in \{1, 2\}$)	1.29	0.249*	(0.098)	99
Number of employees	1.43	0.006	(0.240)	31
Delivered quantity (litres)	389.3	-36.26	(40.69)	99
Delivered to other MCC (litres)	332.7	8.67	(64.76)	47

Note: Balance check for the 99 traders enrolled in Experiment 2 (47 treatment, 52 control). The top panel reports pre-treatment daily-submission outcomes (averages over the period up to 1 November 2025), estimated with MCC fixed effects and trader-clustered standard errors. The bottom panel reports trader characteristics measured at the trader endline survey but plausibly pre-determined at the time of randomization (employment structure and delivered quantities), with MCC fixed effects and heteroskedasticity-robust standard errors. Transport-mode dummies are omitted because all traders use the same mode (motorcycle). Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table A11: Experiment 2 farmer balance

	ctrl	treat	SE	N
Total production (litres/day)	67.32	-2.76	(7.02)	422
Total sold (litres/day)	55.82	-2.69	(6.41)	422
Sold to recruiting trader (litres/day)	49.91	-0.824	(5.85)	422
Years linked to trader	5.04	0.005	(0.102)	422

Note: Balance check for the 422 farmers in the Experiment 2 farmer follow-up survey (197 linked to treated traders, 225 to control traders). Farmers were recruited at endline through traders; no separate baseline survey exists for this sample. Outcomes shown are plausibly pre-determined at the time of trader randomization: farmer-reported total daily production and sales, quantity sold to the recruiting trader, and years linked to that trader. Estimates use MCC fixed effects with trader-clustered standard errors. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table A12: Treatment effects on sales-channel outcomes

	ctrl	treat	SE	N
Sold to MCC (wet season)	0.789	-0.072^{+}	(0.038)	2058
Sold to MCC (dry season)	0.775	-0.066^{+}	(0.037)	2058
Sold to MCC (last 7 days)	0.624	-0.075	(0.047)	2058
Price received (wet season, UGX/L)	992.0	-0.319	(13.11)	1994
Price received (dry season, UGX/L)	1222.4	6.21	(13.38)	1958
Anderson index of sales-channel outcomes	0.055	-0.045	(0.061)	1929

Note: Estimated treatment effects on sales-channel outcomes for the Experiment 1 farmer analysis sample. The first three rows show the probability that a farmer sold milk to the MCC during the indicated period; the next two rows show the average price received per liter conditional on having sold during the indicated period; the last row reports an Anderson index of these channel outcomes. Standard errors are CR2 cluster-robust at the catchment level. The first three rows speak directly to whether treatment differentially affects selection into the price-conditional analysis sample of Table; the marginally negative coefficients indicate mild selection, while the price rows (estimated conditional on having sold) show no statistically detectable treatment effect on prices received. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table A13: Contamination robustness: full sample vs. clean subsample

	Full sample			Clean sample		
	treat	SE	N	treat	SE	N
Panel A: Milk samples (supervised endline testing)						
Butterfat (%)	0.114^{**}	(0.036)	2518	0.104^{*}	(0.036)	2227
SNF (%)	0.081	(0.050)	2518	0.098^{+}	(0.053)	2227
Added water (%)	-0.488^{*}	(0.213)	2518	-0.549^{+}	(0.258)	2227
Protein (%)	0.029	(0.022)	2518	0.047^{*}	(0.021)	2227
CLR	0.435^{*}	(0.197)	2518	0.498^{*}	(0.205)	2227
Quality index	0.209^{**}	(0.054)	2518	0.218^{**}	(0.056)	2227
Panel B: MCC-level outcomes						
Tests incoming milk	0.530^{**}	(0.081)	120	0.751^{**}	(0.068)	106
Tests outgoing milk	0.288^{**}	(0.089)	120	0.403^{**}	(0.086)	106
Farm-gate price	-15.21	(16.36)	115	-6.41	(16.95)	101
Buyer price	3.44	(19.22)	108	2.10	(19.09)	95
Pays quality premium	-0.032	(0.066)	119	0.083	(0.057)	105
Receives quality premium	0.031	(0.071)	119	0.055	(0.073)	105
Uptake index	0.174^{+}	(0.096)	115	0.342^{**}	(0.100)	101
Panel C: Farmer-level primary outcomes						
Improved practices index	-0.020	(0.039)	2054	-0.011	(0.041)	1809
Buyer checks quality	0.071	(0.047)	1337	0.133^{**}	(0.043)	1178
Average price received	11.31	(9.94)	1254	10.14	(10.92)	1104
Receives quality bonus	-0.016	(0.023)	1281	-0.006	(0.023)	1142
Primary farmer index	0.080^{+}	(0.047)	1202	0.134^{**}	(0.048)	1070

Note: Side-by-side comparison of treatment effects on the full estimation sample versus the clean subsample, which drops the 14 control MCCs found to host a functioning milk analyzer at endline (and their associated farmers). Panel A covers the supervised milk-sample outcomes from Table, Panel B the MCC-level outcomes from Table, and Panel C the farmer-level primary outcomes from Table. Across all three panels, point estimates retain their sign and most retain or gain statistical significance on the clean sample, consistent with attenuation from informal analyzer transfers between treatment and control MCCs. Standard errors are CR2 cluster-robust at the catchment level. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

ALL IFPRI DISCUSSION PAPERS

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
www.ifpri.org

IFPRI HEADQUARTERS
1201 Eye Street, NW
Washington, DC 20005 USA
Tel.: +1-202-862-5600
Fax: +1-202-862-5606
Email: ifpri@cgiar.org