Incentivizing quality in dairy value chains - experimental evidence from Uganda (PAP)

Richard Ariong[*], Jordan Chamberlin[†], Sarah Wairimu Kariuki[‡], Bjorn Van Campenhout[§]

October 18, 2022

Abstract

Quality of products transacted within value chains, and the preservation of quality throughout the chain, is central to value chain development. In Uganda, we find that there is a clear demand from dairy processors for better quality raw milk and substantial scope for quality improvement at the dairy farmer level, yet a market for quality does not develop, holding back further value chain transformation. In this study, we test two potential reasons why a market for quality does not develop through a field experiment with randomized interventions at different levels of the value chain. At the dairy farmer level, we conjecture that farmers are paying attention to the wrong quality attributes and design a video-based information campaign to point out what the quality parameters are that matter for processors. We also provide them with a small incentive to put what they learned into practice. Midstream, at milk collection centers where milk is bulked and chilled, we install technology that enables for quick and cheap testing of the milk that is brought in. We look at impact of both interventions at both farmer and milk collection center level and consider outcomes such as milk quality, prices received and quantities transacted.

JEL: O13, O17, Q13

Motivation

Quality of products transacted within value chains, and the preservation of quality throughout the chain, is central to value chain development. Working with quality inputs often reduces production costs further down the value

[*]Development Strategy and Governance Division, International Food Policy Research Institute, Kampala, Uganda

[†]CIMMYT, Nairobi, Kenia

[‡]CIMMYT, Nairobi, Kenia

[§]Development Strategy and Governance Division, International Food Policy Research Institute, Leuven, Belgium

chain. Quality inputs and safeguarding quality while processing, storing, and transporting commodities is also important from a food safety perspective. In general, transformation of value chains often coincide with quality upgrading.

Over the past decade, the dairy sub-sector in Uganda has changed dramatically. Particularly in the areas around Mbarara, commonly referred to as the southwestern milk shed, an influx of foreign direct investment has created the preconditions for modern dairy value chains to emerge (Van Campenhout, Minten, and Swinnen, 2021). The area now has an extensive network of milk cooling and collection centers that link smallholder farmers to a cluster of processors. In the dairy value chain, quality is particularly important. Milk quality determines what products can be produced. For instance, for production of cheese, a high fat content is needed and milk needs to be fresh. To extract caseine, freshness is less important, but the protein content needs to be high. Furthermore, it goes without saying that the protection of milk from dirt and contamination is important for food safety, as milk is very unstable.

At the same time, it is surprising that there seems to be no market for quality in the sub-sector. For instance, using recently collected survey data, we find that of a sample of 200 farmers that sold to milk collection centers, only 6 percent indicated that they received a quality premium. From 114 milk collection centers that were included in the survey, we found that only about 18 percent (sometimes) paid a price premium to farmers. At the same time, expert interviews with processors indicate that their main challenge is related to sourcing milk of sufficient quality, pointing out issues related to butter fat content and solid non-fat content of the milk. They also say that the would be willing to pay for it.

When asked about what farmers need to do to increase quality, farmers mainly refer to practices that affect milk sanitation. Most training and extension activities in the area focus on the importance of using proper equipment (stainless steel milk churns as opposed to plastic jerry cans) and simple practices such as washing hands and udders. These technologies and practices do not affect the milk quality attributes that processors seem to care most about. To increase butter fat content and solid non-fat content, it is especially feeding practices that matter.

The above points to at least two problems which constrain the development of a market for quality milk. First, at a technological level, instruments necessary to make the desired quality attributes visible are lacking. Most milk collection centers only engage in rudimentary testing for adulteration (using a gravity based test with a device called a lactometer) and freshness (using the alcohol test). Farmers do not have access to testing equipment. Second, at the knowledge level, farmers do not seem to know what quality parameters are important further downstream the value chain.

In this research, we will test various hypothesis using a randomized control trial with interventions at both the level of the milk collection centers and at the farmer level. At the level of the milk collection center, we work with the Uganda Dairy Development Authority (DDA) to scale up their Quality-Based Milk Payment Scheme (QBMPS) that was piloted by last year in Uganda's SW milkshed. It involves installing lactoscans at milk collection centers that allows testing of individual milk deliveries for quality parameters desired by processors. We want to test what the impact of visualizing these quality attributes at this level is on both farmers and milk collection centers. We then use a split plot design to mix in a second intervention at the level of the farmers. Here, we provide a video-based information treatment where farmers are informed about what quality parameters processors deem important and how they can improve on these parameters.

This document serves as a pre-analysis plan for the study that will be registered in a public repository. It provides background information, outlines hypotheses which will be tested, tools that will be used in the field, power calculations and sample size projections on which sampling is based, outcome variables that will be used to assess impact, and specification that will be estimated. As such, it will provide a useful reference in evaluating the final results of the study (Humphreys, Sanchez de la Sierra, and van der Windt, 2013; Duflo et al., Working Paper).

Related Literature

Our study is related to a large literature. Some of the most recent articles include:

- Rao and Shenoy (2021) explore the effect of collective incentives on group production among rural Indian dairy cooperatives. In a randomized evaluation, they find village-level cooperatives can solve internal collective action problems to improve production quality. However, some village elites decline payments when they cannot control information disclosure. Opting out reflects frictions in allocating surplus within a social network, and suggests some transparency-based efforts to limit elite capture may undermine policy goals.

- Treurniet (2021) uses matching on observable farmer characteristics to study how individual quality incentives provided by private actors can help smallholders to improve milk quality. In the Indonesian dairy value chains they study, individual quality incentives increased the compositional quality of milk quickly after its introduction. Together with physical inputs and training, individual quality incentives also increased the hygienic quality of milk.

- Saenger et al. (2013) use framed field experiment to evaluate the impact of two incentive instruments: a price penalty for low quality and a bonus for consistent high quality milk on farmers' investment in quality-improving inputs among contract farmers in the Vietnamese dairy sector. Statistical analysis suggests that the penalty drives farmers into higher input use, resulting in better output quality. The bonus payment generates even higher quality milk.

Hypotheses and impact pathways

One potential reason why a market for quality does not develop may be related to the fact that milk from individual farmers is poured together, making it hard to track quality. In general, at the start of the cold chain in milk collection centers, only rudimentary testing is done, and equipment to track quality parameters that are most relevant for the development of a market for quality is lacking. Only when milk reaches the processor, these quality parameters are revealed.

In a first hypothesis, we expect that reducing the cost of quality discovery at the level of the milk collection center (such that it is easy to accurately determine the quality of each individual supplier before it is aggregated in milk tanks) will increase outcomes at that level for several reasons. For instance, it will enable collection centers to turn down suppliers with low quality, which should increase the overall quality of milk aggregated. When milk collection centers are able to independently assess the quality of the milk, they may actively search for processors that are prepared to pay a premium for a particular quality parameter.[1] In addition, accurate information about the quality of the milk may also strengthen the bargaining position of the milk collection center vis-a-vis the buyer. The ability to accurately monitor incoming milk may also enable milk collection centers to engage in product differentiation at an early stage, by for instance using one tank to collect high protein milk destined for casseine extraction and using another tank to collect milk that is high in butter fat, to supply to a cheese maker.

In a second hypothesis, we also expect that dairy farmers will benefit from this intervention at the level of the milk collection centers. Making quality visible midstream should enable milk collection centers to reward farmers for supplying superior milk and increase the overall quality of the milk that the collection center aggregates. If dairy farmers know that the milk collection center has the equipment to test milk at a reasonable cost, farmers may also demand milk collection centers to test their milk in case there is discussion related to the quality.

Another potential reason why a market for quality does not develop may be related to the fact that farmers do not have adequate knowledge about what is meant by milk quality. In particular, farmers seem to focus most on food safety related quality aspects of milk, and less on the compositional aspect. As a result, even when the technology to assess quality is available, farmers may not be able to improve without additional knowledge on what parameters to improve upon. Furthermore, it may be that farmers do not have a good understanding of how these compositional parameters can be affected.[2] A third hypothesis is thus

[1] As mentioned earlier, milk quality determines what products can be produced. If the milk collection center discovers their milk has a particularly high butter fat content, it may decide to deliver to a cheese producer who is prepared to pay more for high fat milk than a processor that extracts caseine who is more interested in SNF.

[2] Being a non-rival good, information is generally undersuplied by the private sector. Agricultural extension and advisory services are therefore often organized by governments or non-

that providing information on what the desired milk quality parameters are, and what affects these parameters, increases outcomes for farmers.

In value chains, it is not always clear whether upgrading is driven by push (eg a productivity increasing technological innovation at the farm level) or pull factors (eg in increase in demand due to opening up of export markets). Often, it is a combination of both, and push and pull factors endogenously reinforce each other in a virtuous cycle (Van Campenhout, Minten, and Swinnen, 2021). In a final hypothesis, we thus also test if making quality visible at the milk collection center level and at the same time providing information on what the desired milk quality parameters are increases outcomes for farmers.

Experimental design

The field experiment consists of two cross-randomized interventions that are implemented at different levels. Outcomes may be measures at different levels. The design is illustrated in Figure 1, which provides a stylized representation of the dairy value chain. We randomly allocate quality testing equipment to a random subset of milk collection centers (MCCs), while another random subset of milk collection centers functions as the control group for this treatment. In the catchment area of each milk collection center, we then take a sample of dairy farmers, stratifying the sample on whether the farmer is an active supplier to the milk collection center or not. In this sample, we then randomly assign half of the farmers to the information treatment (blocking on whether the farmer is an active supplier to the milk collection center or not).

With this design, we can then test if the intervention at the milk collection center improved outcomes for milk collection centers. We can also test if the intervention at the milk collection center affects outcomes at the farmer level by comparing outcomes of the farmers in catchment areas of treated Milk Collection Centers (MCCs) to outcomes of farmers in catchment areas of control MCCs. The intervention at the farmer level can only be evaluated at the farmer level. At the level of the farmers, we can also look at the interaction between the two treatments by looking at outcomes of farmers that received the information treatment in catchment areas of milk collection centers that also received a lactoscan in relation to outcomes of farmers that are differently exposed to the treatments.

In sum, and in reference to the equation we will estimate in the next section, the four main hypotheses that we will test with this design are:

- Hypothesis 1: making quality visible at the MCC level increases outcomes for the milk collection centers ($\beta_{H1} > 0$).
- Hypothesis 2: making quality visible at the MCC level increases outcomes for the farmers in the catchment areas of these MCCs ($\beta_{H2} > 0$).

governmental organizations who tend to prioritize food safety concerns over profitability. As a result, farmers are mostly trained on how to maintain milk sanitary standards and less on ways to improve quality in terms of butter fat and Solid Non-Fat.

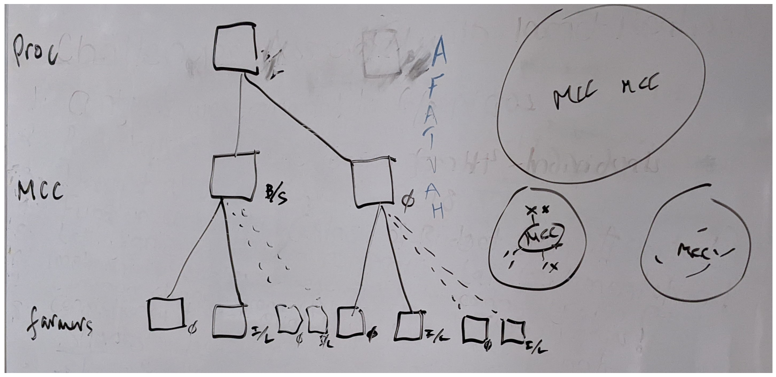

Figure 1: Design

- Hypothesis 3: providing information on what the desired milk quality parameters are and what affects this parameter increases outcomes for farmers ($\beta_{H3} > 0$).
- Hypothesis 4: making quality visible at the MCC level and providing information on what the desired milk quality parameters are to farmers increases outcomes for farmers ($\beta_{H4} > 0$).

Additional research questions, based on the stratification, tests for differences in average treatment effects between farmers that are connected to milk collection centers versus those that are not. Testing for this treatment heterogeneity allows us to explore if the interventions only strengthen existing value chains or whether they can also draw in actors from informal value chains.

- Does the MCC level intervention affect farmers that are already connected to the milk collection center differently than farmers that are not already connected to an MCC ($\beta_{H2C} = \beta_{H2}$).
- Does the information treatment affect farmers that are connected to an MCC differently than farmers that are not connected to an MCC ($\beta_{H3C} = \beta_{H3}$)
- Does the combined treatment (making quality visible at the MCC level and providing farmers with information on the desired quality dimension) affect farmers that are connected to an MCC differently than farmers that are not connected to an MCC ($\beta_{H4C} = \beta_{H4}$)

Interventions

To make relevant quality parameters visible at the level of the milk collection centers, we focus on a technological intervention. In close collaboration with DDA, we install digital lactoscans at a random sample of milk collection centers.

These can be used to test milk samples of individual farmers or traders that supply to the milk collection centers to establish quality of incoming milk, as well as to test samples from the milk tankers when milk is picked up by traders or processors.

To provide information to dairy farmers on the parameters and characteristics that processors are looking for and how farmers can produce milk that adheres to these standards, we use a short engaging video that demonstrates the inputs and practices that can be used to increase milk quality. The use of video has been found to increase technology adoption in different settings, although the effectiveness also depends on a range of design attributes (Spielman et al., 2021). The ability to depict role models in videos seems important to increase both aspirations of the person targeted, as well as creating an enabling environment for adoption in that it may challenge world views and stereotypical thinking (Riley, 2019; Lecoutere, Spielman, and Van Campenhout, 2020).

To design the video based extension intervention, we first identified the top five practices and inputs that are known to raise butter fat and Solid Non Fats in milk. This was done through consultations of experts. We found the top 5 practices and inputs were: selection of breed and genetic potential, selection of grasses for high-quality forage, best practice in silage and hay making, correct mixing and dosage of feed, and feed supplements like Methionine and Lysine. To make the information intervention more actionable, we also provide farmers with some free inputs (feed supplements and/or seed for eg Napier grass).

Estimation and inference

We will estimate two equations using Ordinary Least Squares. One equation is at level of the milk collection centers, the second equation is at the level of the dairy farmers.

Denote milk collection centers by m, running from 1 to M. T_m is a treatment indicator at the MCC level that is one if the MCC (in who's catchment area the farmer resides) was allocated to the lactoscan treatment. y_m is the outcome at the level of the milk collection center you want to estimate the treatment effect for and ε_m is an error term.

$$y_m = \alpha + \beta_{H1}.T_m + \varepsilon_m \tag{1}$$

The parameter of interest in this equation is β_{H1}, which tests Hypothesis 1.

The second equation is at the individual level. Here, T_i is a treatment indicator at the farmer level that is one if the farmer was allocated to the information treatment (with i indicating the farmer running from 1 to I). $C_{i,m}$ is an indicator variable at the farmer level that is one if the farmer i is connected to MCC m and zero otherwise. $y_{i,m}$ is the outcome of interest at the level of the individual farmer living in the catchment area of milk collection center m and $\varepsilon_{i,m}$ is an error term (which may be correlated within catchment area).

$$y_{i,m} = \alpha + \alpha_C C_{i,m} + \beta_{H2}.T_m + \beta_{H3} T_i + \beta_{H4} T_i.T_m$$
$$+ \beta_{H2C}.T_m.C_{i,m} + \beta_{H3C} T_i.C_{i,m} + \beta_{H4C} T_i.T_m.C_{i,m} + \varepsilon_{i,m} \quad (2)$$

Standard errors in equation 2 are clustered at the milk collection level. The parameter of interest in this equation is β_{H2}, which tests Hypothesis 2, β_{H3}, which tests Hypothesis 3 and β_{H4}, which tests for the interaction effect. We also add a full set of interactions with the connection indicator to look at treatment heterogeneity.

Factorial designs have recently been criticized for the proliferation of underpowered studies and replication failure (Muralidharan, Romero, and Wüthrich, 2019). While in the next section we will run power calculations based on models with a complete set of interactions (equation 2), we may still want to try boosting power by pooling observations across the orthogonal treatment in the event that we find a treatment effect that appears smaller than the minimal detectable effect size that we assumed during power calculations. To do so, we will consider the orthogonal treatment as a co-variate we adjust for, and interact the treatment variable with the demeaned orthogonal treatment. This give a more robust version of the treatment estimate that corresponds to the coefficient estimate of the treatment of interest after dropping the interaction with orthogonal treatment.

Finally, to increase precision of the estimates of the treatment effects, one often includes baseline outcomes as controls and estimate ANCOVA models (Duflo, Glennerster, and Kremer, 2007). However, recent research has shown that with covariate adjustment, 1) the conventional OLS standard error estimator is inconsistent and 2) adjustment can sometimes hurt precision asymptotically. The former can be fixed by using a sandwich estimator; we will use HC3 sandwich estimator which performs better in small samples. The latter can be fixed by including baseline outcomes as deviations from their mean value and interacting them will the treatment variable (Lin, 2013). This will lead to another full set of interactions in equation 2.

We will use simulation to account for multiple comparisons. Simulation methods provide a flexible and intuitive way to think about multiple hypothesis testing. It accommodates the extent to which the multiple comparisons are correlated with one another and allows us to integrate design specific elements such as blocking and multiple arms. This leads to a study-specific correction that will generally be more powerful than other methods to control the FWER. In particular, to determine target p-value cutoffs, we use the family-wise sharp null of no effect for any unit on any dependent variable and for any hypothesis. We will also combine primary outcomes into an index following Anderson (2008), which also guards against the dangers of multiple comparisons (See Section).

Power calculations

We also use simulation to determine sample size. The primary outcome variable that we use in our statistical power calculations is the price of milk.

We start at the level of the milk collection center and assume that at this level, the price at which milk collection centers sell their aggregated milk is normally distributed with mean 1000 UGX per liter and standard deviation of 50 (which is half of what we will assume at the farmer level). From these N observations (with N denoting the number of milk collection centers recruited for our study and hence the first key variable to be determined by the power calculations) we then generate N times n observations. These are the n dairy farmers that are located in the catchment areas of the N milk collection centers. The outcome variable at this level, prices that farmers obtain from milk collection centers, are generated again as random normal, but with the mean the value that was drawn for the MCC the n farmers are connected to, and with a slightly higher standard deviation (100 — since, as the milk is not aggregated yet, extreme values are not yet averaged out). This procedure gives us a total sample with N prices at the MCC level and $N.n$ prices at the farmer level, the latter being clustered at the MCC catchment area level by design.

We assume that the intervention at the level of the milk collection centers leads to an increase in the price of UGX30 per liter. This seems reasonable in light of the fact that processors told that they either pay a 10 percent premium for quality milk, or UGX100 per liter. However, as we assume a pretty narrow distribution of prices, even though this effect is only a 3 percent increase, this is considered a medium to large effect according to Cohen's D. At the level of the farmers, for the intervention at the MCC level, we expect an effect size of UGX40. While this represents a 4.4 percent increase, the larger variance at this level means that according to Cohen's D, this effect is considered small to medium. Finally, at the level of the farmers, the individual level randomization of the information treatment intervention allows us to estimate small effects. For our power simulation, we assumed and effect size of UGX25, which corresponds to a small effect according to Cohen's D. For the interaction, we assume a large effect (UGX50 per liter).

We calculate power for the joint test that the three hypotheses are true at the 5 percent significance level. To do so, we run the exact two regressions from Section and run 1000 simulations for each n*N combination. For each n*N combination, we calculate the share of simulations at which all coefficients of interest in Equations 1 and 2 (β_{H1} to β_{H4}) are significant at the 5 percent level to determine power.

Results of the simulation are summarized in Figure 2. Instead of the usual power curves that plot power against sample size, we obtain a power plane as we determine both the number of clusters (between 100 and 130 MCCs) and the number of farmers per cluster (between 10 and 40 farmers). Power is measured on the z axis and is the proportion of cases (out of the 1000 simulations) in which all three coefficients were found significant at $p<0.05$.

The figure, which can be found as an interactive figure here, shows the trade-

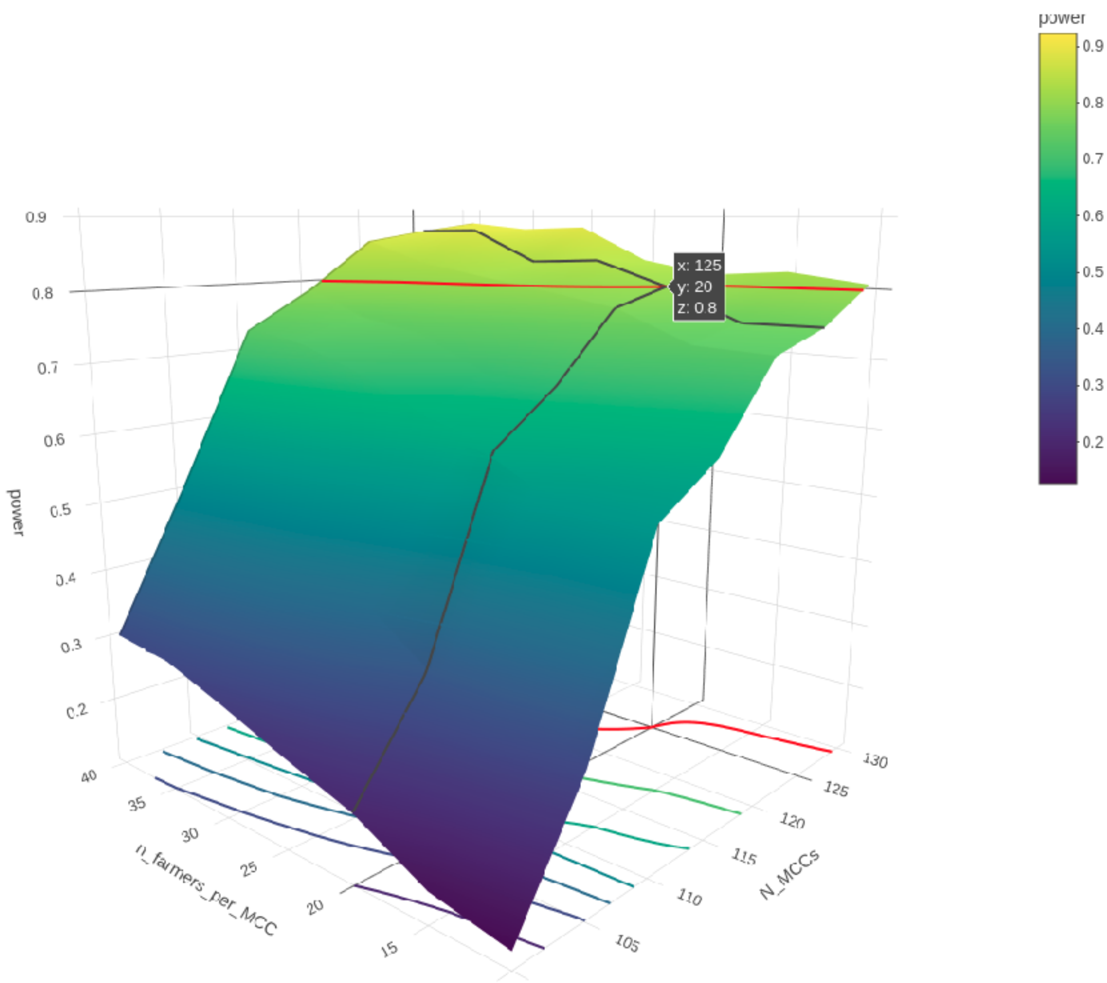

Figure 2: Power plane

off between more clusters and more individuals per cluster. With about 125 MCCs and 20 farmers per cluster we find power just above .80. This corresponds to a sample of 2500. Note that the requirement to detect minimum effect sizes for all three hypotheses simultaneously is very strict. For instance, if we require only one hypothesis to be significant, we obtain power of .99 for a sample with 125 MCCs and 20 farmers. Similarly, if we consider each hypothesis separately, we get power levels of .87 for the MCC level intervention with outcome at the MCC level, .93 for the MCC level intervention with outcome at the farmer level, and .94 for the farmer level intervention with outcome at the farmer level. For the interaction effect, we obtain power of .99.

Timeline

We plan to collect baseline information in November-December 2022. During that time, we will also implement the intervention at the level of the farmer. Immediately after baseline data collection, we will also start installing lactoscans in the selected milk collection centers. This is expected to take about three months, such that all lactoscans are install towards the end of March 2023. Midline data will be collected about half a year after the last lactoscan was installed, so this will be in September-October 2023. Endline data will be collected one year after the intervention, which is April 2024.

Data collection and outcomes of interest

Sampling

We start from a list of registered milk collection centers that was obtained from the Dairy Development Authority. From this list, we randomly selected 130 milk collection centers, half of which were assigned to the treatment group using a computer algorithm. We then travel to these 130 milk collection centers and use systematic sampling to get a sample of 10 farmers that are delivering to the MCC. In particular, we will visit the MCC early in the morning and get an estimate of the expected number of farmers that will visit during the course of the day. This will be used to determine the interval at which farmers will be picked to participate in the study. These farmers will be interviewed at home the next day. We will also contact the immediate neighbor of this farmer and, if this farmer is not delivering to the MCC, the farmer will also be included in the study.

Demonstrating balance

During baseline data collection, we will collect information on the following 10 variables at each level to demonstrate balance (although we may only report a subset - indicated with a star - in the paper for space considerations) At the milk collection center level, we will collect the following characteristics:

1. Is this milk collection center (part of a) cooperative? (yes/no)*
2. Number of people employed (full-time) at this MCC? (number)
3. Number of farmers/traders that supply on an average day during the rainy season. (number)
4. Total Capacity of MCC (in liters)*
5. Capacity use during dry season (percentage)
6. Does the MCC pay a premium for quality (yes always or yes, sometimes = 1)*
7. Years Experience in MCC*
8. Number of milk cans owned by the MCC
9. Supplies credit/loans to cooperative members and supplying farmers? (yes=1)
10. Facilitates supply of acaracides to cooperative members and supplying farmers? (yes=1)*

At the level of the farmer household, we will collect information on the following 10 characteristics:

1. Household Members (number)
2. Household Head Age* (years)
3. Current Total herd size (cows+heifers+calves) (number)*
4. Number of improved animals in total herd (share) *
5. Liters Produced Total Per Day (average during rainy season) (liters)
6. Liters milk sold per day (on average in the rainy season) (liters)*
7. Normally during the rainy season sells most of its milk to a milk collection center? (yes=1)
8. Uses only steel can/bucket during sales transactions? (yes=1)
9. Member of dairy cooperative? (yes=1)
10. What is your average monthly expense (UGX) on chemical purchases to fight ticks (acaracides)? (average during rainy season)*

Primary outcomes

We define five primary outcomes at each level. These five primary outcomes will be combined in a covariance weighted index to assess overall impact at that level following (Anderson, 2008). As dairy is a continuous activity, we need to define a time frame for measurement. We will use the last full week before the interview. The five primary outcomes at MCC level are:

1. average milk quality level of milk sold. This will be sampled from the milk tanks, and based on an index of different quality parameters (at least butter fat content and SNF).

2. average prices at which milk was bought from farmers (during last 7 days)

3. volumes collected in last 7 days

4. sold to top 5 processors (Pearl, Amos, Lakeside, GBK, Vital tomosi) (in last 7 days)

5. price at which milk was sold (in last 7 days)

Outcomes of interest at farmer level, measured in the last seven days:

1. Milk quality (butter fat content and SNF)

2. Production investment and management (based on index of five recommended practices to improve milk quality)

3. Volumes sold (liters during last week)

4. Sold to milk collection center during last week? (1=yes)

5. Price received for milk sold (inclusive of any quality premium that may have been obtained) (average during last week, UGX per liter)

Secondary outcomes

Secondary outcomes at the milk collection center level include:

1. local sales - previous research found that milk collection centers are also important for local milk supply, often doubling as milk shops. Does the intervention crowd out the local market?

2. reason for selling to buyer (in particular if the buyer pays premium for quality, but also payment modalities)

3. Impact pathway: did MCC measure quality of aggregated milk before selling? In particular butter fat and SNF using a lactoscan? What equipment was used?

4. Who decided on the price? buyer made offer and MCC accepted, MCC made offer and buyer accepted, negotiation — use likert scale slider to get an idea of power balance.

5. Did the buyer pay a quality premium? What was it based on? What is the quality premium?
6. Does the MCC pay a quality premium to suppliers? What was it based on? What is the quality premium?
7. Does market for quality lead to additional investment in quality preservation - milk cans, etc
8. Does the development of a market for quality lead to more formalization (eg written contracts) between farmer and MCC? Between MCC and processor?
9. Changes in mid-stream service provision: Does the MCC provide services related to artificial insemination? Transport? Access to acaracides? Training on milk sanitation? Training on feeding practices?
10. Information on lactoscan use (for ITT-TOT analysis).

Secondary outcomes at the farmer level include:

1. Home consumption of dairy products (liters, in what form, and who consumes diary products) - test if the development of a market for quality milk crowds out animal sourced food intake within the family.
2. Reason for selling to buyer (in particular pays premium for quality, payment modalities,...)
3. Test if intervention leads to quality based market segmentation (with less rejection and more instances of lowering of price when farmer supplies substandard milk)
4. Does the buyer pay for higher quality milk.
5. Buyer checks for quality during last transaction (lactoscan, lactometer, alcohol test).
6. Number of dairy animals (improved/local) - does a market for quality lead to technology adoption for intensification? Is this stronger for the subgroup of farmers that receives the training video, where we explicitly mention that genetics also affect quality parameters?
7. Feed and pasture management - a more detailed analysis than the composite primary outcome 2 at the farmer level. This includes changes in grazing system (paddocking, free range, mixed or zero grazing) and use of different dairy feed types (hay, silage, improved forages, commercial feeds like (brewers) bran, salt and mineral blocks, multivitamin). We will differentiate between practices in the rainy season and the dry season.
8. Price of dairy animals (improved/local) - test if the development of a market for quality has an impact on the price of animals.

9. Gendered decision making outcomes - test if the development of a market for milk impacts who within the household makes the decisions to sell to a particular buyer.

10. Does the development of a market for quality lead to more formalization and less relational contracting?

11. Does the intervention also increases milk sanitation (use of milk cans)?

12. Does the intervention leads to changes in bargaining power? farmer made offer and MCC accepted, MCC made offer and farmer accepted, negotiation — use likert scale slider to get an idea of power balance.

13. Gendered labour outcomes (milking, marketing, feeding and herding or cleaning)

14. Does the intervention affect home processing? Does this have gendered effects?

Ethical clearance

This research received clearance form Makerere's School of Social Sciences Research Ethics Committee () as well as from IFPRI IRB (). The research was also registered at the Ugandan National Commission for Science and Technology ().

Transparency and replicability

To maximize transparency and allow for replicabiliy, we use the following stratgies:

- pre-analysis plan: the current document provides an ex-ante step-by-step plan setting out the hypothesis we will test, the intervention we will implement to test these hypothsis, the data tha will be collected and specifications we will run to bring the hypotheses to the data. This pre-analysis plan will be pre-registered at the AEA RCT registry.

- revision control: the entire project will be under revision control (that is time stamped track changes) and committed regularly to a public repository (github).

- mock report: After baseline data is collected, a pre-registered report will be produced and added to the AEA RCT registry and GitHub. This report will differ from the pre-analyisis plan in that it already has the tables filled with simulated data (drawn from the baseline). The idea is that after the endline, only minimal changes are necessary (basically connecting a different dataset) to obtain the final result, further reducing the opportunity of specification search.

Funding

This research is part of the OneCG initiative on Rethinking Food Markets which is funded by a consortium of donors. The project builds on previous work on dairy value chains in Uganda that was funded by the Belgian Public Service for Foreign Affairs, Foreign Trade and Development Cooperation, grant number 302015.002.002 (bridge funding). This grant also funded part of the fieldwork for baseline data collection.

References

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.

Duflo, E., R. Glennerster, and M. Kremer. 2007. "Using randomization in development economics research: A toolkit." *Handbook of development economics* 4: 3895–3962.

Duflo, E., A. Banerjee, A. Finkelstein, L. F. Katz, B. A. Olken, and A. Sautmann. Working Paper. "In Praise of Moderation: Suggestions for the Scope and Use of Pre-Analysis Plans for RCTs in Economics."

Humphreys, M., R. Sanchez de la Sierra, and P. van der Windt. 2013. "Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration." *Political Analysis* 21 (1): 1–20.

Lecoutere, E., D. J. Spielman, and B. Van Campenhout. 2020. *Women's empowerment, agricultural extension, and digitalization: Disentangling information and role-model effects in rural Uganda*. Intl Food Policy Res Inst.

Lin, W. 2013. "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique." *The Annals of Applied Statistics* 7 (1): 295–318.

Muralidharan, K., M. Romero, and K. Wüthrich. 2019. *Factorial designs, model selection, and (incorrect) inference in randomized experiments*. Tech. rep., National Bureau of Economic Research.

Rao, M. and A. Shenoy. 2021. "Got (clean) milk? governance, incentives, and collective action in Indian dairy cooperatives."

Riley, E. 2019. "Role models in movies: the impact of Queen of Katwe on students' educational attainment." *The Review of Economics and Statistics* 1–48.

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.

Spielman, D., E. Lecoutere, S. Makhija, and B. Van Campenhout. 2021. "Information and communications technology (ICT) and agricultural extension in developing countries." *Annual Review of Resource Economics* 13: 177–201.

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., 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.