The Potency of Quality Incentives: Evidence from the Indonesian Dairy Value Chain

Mark Treurniet

Misalignment of quality incentives along value chains may limit quality and therefore function as a barrier to smallholder participation in modern value chains. This article uses survey and administrative data to study how individual quality incentives provided by private actors can help smallholders to improve milk quality. By matching farmers on baseline characteristics, I find that 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. Decreasing hygienic quality over time by treated farmers suggests that the impact of the intervention decreased over time.

Key words: Agricultural technology, market incentives, quality.

JEL codes: O13, Q12, Q13.

Although increasingly stringent food quality standards benefit consumers in many ways, they can also function as a barrier to entry on the supply side. For smallholder producers in developing countries in particular, such demanding standards can make participation in modern value changes difficult. To be able to meet quality standards, farmers need to have access to a suitable technology, know how to use the technology, and be able to finance the technology (Foster and Rosenzweig 2010; Jack 2013).

But even if smallholders are able to adopt the right technologies, they are often insufficiently incentivized to do so. As smallholders by definition produce small amounts, their production tends to be aggregated before further being marketed and processed. To reduce transaction costs, quality tests are often conducted only at the aggregated level so that individual quality cannot be rewarded, and investment in quality is essentially a public good. Examples are widespread and are documented for wheat in Ethiopia (Abate et al. 2019), maize in Kenya (Hoffmann and Jones 2018; Hoffmann et al. 2020), groundnuts in Ghana (Magnan et al. 2019), and onions in Senegal (Bernard et al. 2017).

However, new and cheaper technologies to measure quality and administer incentives provide the potential to align incentives along value chains by reforming small-scale food transactions and might help smallholder farmers to increase quality and keep up with increasing quality standards (Hoffmann, Moser, and Saak 2019).

Mark Treurniet is a postdoctoral researcher at Wageningen University & Research. The author thanks the editor, Travis Lybbert, four anonymous reviewers, Erwin Bulte, Jos Bijman, Stefano Caria, Kasper Hettinga, Marlene Roefs, and participants at the DEC PhD Club at Wageningen University and the MTID-PIM Brown Bag Seminar at IFPRI for helpful suggestions and comments. Marlene Roefs identified the opportunity to study the impact of the intervention under scope and initiated the baseline data collection together with Anja Wolsky. Andre Daud and Safitri Tejaningsih provided great support in the field. The cooperative described in this article provided access to the administrative data used in this article. Financial support for this project has been received from IFPRI through the CGIAR Research Program on Policies, Institutions and Markets.
Correspondence to be sent to: mark.treurniet@wur.nl

Amer. J. Agr. Econ. 00(00): 1–18; doi:10.1111/ajae.12176

This article provides evidence that the alignment of incentives along the value chain can lead to substantial smallholder production responses. I study the effects of an intervention in the Indonesian dairy value chain that introduced individual quality incentives for cooperative members. Before the intervention, farmers were paid based on the average quality delivered by a group of farmers. In an effort to increase the hygienic quality of milk, part of the cooperative's milk collection points (MCPs) were upgraded, associated farmers were trained and an individual quality incentive system was introduced. After the upgrade, prices were based on the quality delivered by the individual farmer using the same price function as before (except that some extra bonuses were introduced for good hygienic quality of milk).

I use a combination of survey data and cooperative administrative data to match farmers that were offered individual quality incentives to similar farmers that were not. A combination of cooperative administrative data and additional quality measures is then used to study the impact on different quality dimensions. The panel structure of the administrative data allows me to study effects over time.

Results show that the intervention led to improved milk quality. Although the upgrade of facilities and the training focused on the hygienic quality of milk, the individual price incentive scheme also focused on compositional quality, which refers to the chemical substituents of milk, like fat, protein, and lactose. To study the isolated effect of individual quality incentives, I estimate the effect of the intervention on compositional quality. I find a positive impact on compositional quality, suggesting that individual incentives can be sufficient to improve food quality, even if they are not combined with agricultural training and physical inputs. Together with the physical upgrade and intervention trainings, individual quality incentives also increased hygienic quality. Decreasing hygienic quality at upgraded MCPs suggest that intervention impacts decreased over time as the physical and human capital inputs delivered by the intervention decayed.

This article contributes to a small, but emerging, body of empirical literature on how smallholders respond to market incentives. Generally, incentives are found to matter in the context of agriculture in developing countries (Casaburi and Macchiavello 2015; Bernard et al. 2019; Burchardi et al. 2019). Quality incentives, however, are a special case, as they require costly measurement of otherwise often unobservable quality. Depending on who measures quality, and the transparency of the measurement, this introduces a unique form of asymmetric information in which the buyer has more information about product characteristics than the producer (Saenger, Torero, and Qaim 2014; Abate and Tanguy 2017).

A number of studies consider impacts of offering quality incentives where quality is measured by a research body or a third party. In most of them, quality is assessed by research bodies and incentives are financed by research funds. Saenger et al. (2013) use a lab-in-the-field experiment to study the effect of offering quality incentives to dairy farmers in Vietnam and found that both a higher penalty for low quality milk and an extra bonus for high quality increase the use of inputs. Three field experiments found that offering price premiums for conforming to local aflatoxin standards to groundnut farmers in Ghana and maize farmers in Kenya increases the adoption of aflatoxin preventing technologies and decreases aflatoxin contamination (Hoffmann and Jones 2018; Magnan et al. 2019; Hoffmann et al. 2020). In one study, by Bernard et al. (2017), quality is assessed by a third party and quality incentives are provided by private actors. Starting from the premise that quality incentives can only be offered if quality is measured, the authors randomly provided villages with information on the introduction of scales and quality labelling on local onion markets in Senegal, which allowed sales to be based on weight and quality rather than volume alone. They found that their information treatment increased the use of quality enhancing inputs and quality. Only after the announced scales and labelling were actually introduced with some delay, treated farmers also received significantly higher prices and earned higher incomes. Yet, all existing articles consider quality measurements by research bodies or external parties, which may be trusted more but may also involve high transaction costs. To the best of my knowledge, this is the first article that shows that quality incentives can also work on the basis of measurements done by the buyer.

On top of that, existing field studies consider effects during at most one agricultural season, which are therefore driven by farmer's expectations on the production function and quality measurement. By exception, Magnan et al. (2019) study effects during two agricultural seasons to allow some trust to grow after the first season. In contrast, this article studies impacts over multiple years, during which farmers can monthly update their information after receiving signals from the incentive system.

Understanding the impact of food quality incentives is important for several reasons. From an economic perspective, it helps to understand how a simple market innovation can help to create value. From a development perspective, it helps to understand how smallholders can be incentivized to comply with food standards, which may be essential to guarantee that smallholders can benefit from access to output markets. From a public health perspective, to the extent that higher quality contributes to safer and more nutritious food, it provides insights in how the market can deliver this.

The remainder of this article is structured as follows. I first model various mechanisms in which quality can be improved and then discuss the empirical setting used to test model predictions. After discussing the data, econometrics and results will be discussed by quality dimension. I close with a discussion.

Model

In this section, I will present a simple model that captures several important mechanisms in which quality can be improved. In the model, farmers exert costly effort to produce quality. They are member of a so-called payment group, and quality is measured at the level of this group. Based on the group measure, farmers receive an incentive. I will use the model to derive comparative statics for quality with respect to changes in the group size, the quality premium, and the cost of effort.

Because most of the fixed investments needed for quality improvement are provided by the intervention studied, the remaining costs to improve quality are mostly variable and proportional to quantity. I therefore abstract from quantity effects and consider the utility per unit produced. The utility of a farmer i per kg produced is increasing in the price and decreasing in the costs of production:

$$(1) \quad u = \alpha + \beta q^* - \theta c(e_i)$$

where α is the basic price, β is the quality premium, q^* is the observed quality, θ is a scaling parameter for the cost of effort, and $c(e_i)$ is the cost of effort as a function of effort e_i. The cost of effort can be affected by improving physical capital, increasing knowledge and focusing scarce bandwidth (mental capacity and executive control) to quality issues (Schilbach, Schofield, and Mullainathan 2016; Treurniet 2020).

Observed quality q^* is the weighted mean of the quality of the members of the payment group to which farmer i belongs:

$$(2) \quad q^* = \sum\nolimits_{j \in G} q_j \omega_j$$

where q_j is the quality produced by farmer j, and ω_j is the share of the production of farmer j in the total production of payment group G with $\omega_j \geq 0$ and $\sum_{j \in G} \omega_j = 1$.

Quality q_i is the result of effort:

$$(3) \quad q_i = e_i$$

Finally, the cost of effort function is both increasing and convex in effort e_i.

Maximizing (1) with respect to e_i yields the following first-order condition, which uniquely defines optimal effort e_i^* as long as the farmer produces milk:

$$(4) \quad c'\left(e_i^*\right) = \frac{\omega_i \beta}{\theta}$$

Effort and quality are thus increasing in the ratio $\frac{\omega_i \beta}{\theta}$, which can be interpreted as the size of the effective quality premium relative to the cost of effort parameter. Effort will increase if the share ω_i of farmer's milk in the total group sample increases, the quality premium β increases, or the cost of effort parameter θ decreases.

This results in three hypotheses that will be tested this study:

1. If individual quality incentives replace group incentives, then the share ω_i of farmer's milk in his relevant sample increases to 1, and quality q_i increases.
2. If the quality premium β increases, then quality q_i increases.
3. If the relevant cost of effort parameter θ decreases, then quality q_i increases.

Empirical Setting

To test these three theoretical predictions, I study the effects of an intervention at a local dairy cooperative in a peri-urban setting in Indonesia. About 3,700 dairy farmers deliver almost all of their milk twice a day to one of the cooperative's 31 milk collection points (MCPs).[1] The cooperative aggregates the milk and sells most of it to dairy processing companies. In addition, the cooperative provides inputs, extension services, financial services, and access to a health facility.

The cooperative faces several challenges in sourcing high quality milk from its members. First, producing good hygienic quality is mainly a behavioral issue. To avoid bacterial contamination, producers should clean and dry the cow's teats before milking, throw away the first milk, which contains most bacteria, filter the milk adequately, and use a proper and clean milk can. Farmers are said to be aware of these principles, but the challenge is to have these implemented.

Second, compositional quality refers to the mix of chemical constituents and is regarded to be higher if the milk contains less water and more solids, like fat, protein, and lactose. In general, the compositional quality of milk mainly depends on breed, cow health, feed, and added water. Although compositional quality varies with cow breeds, broad replacement of cows by another breed is not expected within the scope of this study. Increasing hygienic practices could decrease udder infections and lead to measurable quality improvements over a period of months to a year. Further, most solids are produced when cows are sufficiently fed with an optional combination of easily digestible concentrate and roughage, which takes longer to digest but promotes a well-functioning digestion. In this specific context, where grasses are scarcely available, the solid content can be gradually increased over a period of weeks to months by increasing roughage intake. Finally, although adulteration with water is said to be rare, the quickest way to increase the solid content of milk is to avoid water to be added accidentally or purposefully.

To incentivize members to produce good quality milk, the cooperative measures the quality delivered by groups of 3–8 farmers.[2] Farmers receive a higher price if the quality of their so called payment group is higher. More specifically, farmers received an incentive that linearly increased with the group's compositional quality as measured by the percentage of total solids; penalties if the freezing point of the group milk exceeded one or two thresholds, indicating that water is added to the milk; and additional bonuses if the number of bacteria in the group milk was below certain thresholds. Figure 1 shows how prices vary, ceteris paribus, as a result of variation in each of these three quality variables during February–December 2018, and thus gives an idea of the strength of the incentives.[3] Though using group samples saves testing costs, farmers may not internalize the full benefits of quality improvement, as part of the benefits end up with fellow payment group members.

In an effort to, among other things, increase the hygienic quality of milk and adhere to the Indonesian National Standard (SNI) of 1M bacterial colony-forming units per milliliter (cfu/mL), a public-private partnership (PPP) was initiated between the cooperative, one of its main buyers, another local dairy cooperative and several supporting non-profit organizations, and granted a subsidy of several million euros.[4] Starting in February 2015, the

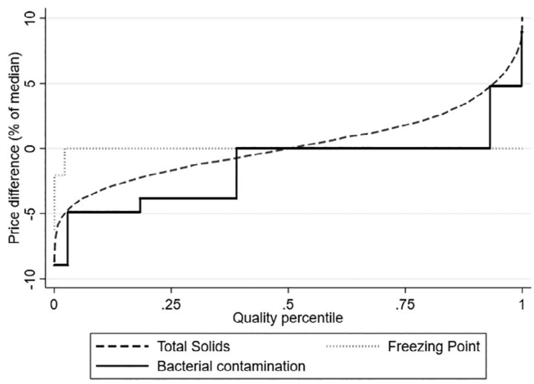


Figure 1. Variation in quality and prices during February–December 2018

[1]Based on cooperative monitoring data for 2014.

[2]Interquartile range in cooperative monitoring data for 2014. The minimum is 1, the median 6, and the maximum 22.

[3]Over time extra bonuses were introduced for low levels of bacterial contamination, so the line for bacterial contamination would be flatter, that is, have fewer steps, in earlier periods.

[4]"Development of Sustainable Dairy Villages in Indonesia" is a project of FrieslandCampina Nederland Holding BV, PT Frisian Flag Indonesia, The Friesian Agro Consultancy BV, Stichting Agriterra, Wageningen UR (Stichting Dienst Landbouwkundig Onderzoek), Koperasi Peternak Sapi Bandung Utara (KPSBU) Jabar, Koperasi Peternak Bandung Selantan (KPBS) Pangalengan and the Ministry of Foreign Affairs of the Netherlands through the Facility for Sustainable Entrepreneurship and Food Security (FDOV). Wageningen University is not part of this PPP.

PPP began to upgrade MCP facilities, train associated farmers, and implement an individual incentive system for these farmers. Over a period of three years, the intervention was implemented at seven MCPs.

First, the MCP facilities were upgraded. A new registration and sampling system was installed, and new cleaning facilities for milk cans were built. Farmers also received a loan to buy a new milk can and a complimentary filter and bucket.

Second, four to eight months before the opening of the new MCP, farmers were invited to attend a socialization meeting in which the upgrading plans were explained. In the following months, farmers attended four to six two-hour trainings on the four hygienic practices discussed above: clean and dry the cow's teats before milking, throw away the first milk, filter the milk adequately, and use a proper and clean milk can. Meanwhile, farmers were visited by extension officers who used checklists to monitor practices, and individual samples were taken to monitor hygienic quality. When more than 95% of the farmers correctly implemented the four hygienic practices, the total plate count (TPC) dropped below 500,000 cfu/mL for all farmers during pre-opening individual testing, and the renovation of the MCP building was completed, the upgraded MCP opened.

Third, starting from the opening of the upgraded MCP, farmers received incentives based on their individual quality for both hygienic and compositional dimensions. In addition, the number of bacteria was more precisely measured. Over time, extra bonuses were introduced for low levels of bacterial contamination at both upgraded and non-upgraded MCPs. Importantly, the price structure itself did not differ between upgraded and non-upgraded MCPs.

Because compositional milk quality was not the focus of the intervention, compositional quality was not addressed in trainings and the size of premiums for compositional quality was not changed either. Further, except through long-run effects on cow health, the promoted hygienic practices are unlikely to have substantially affected compositional quality.[5] As a side effect of the intervention, however, farmers also received individual incentives for compositional quality. I use this dimension of quality to test the first hypothesis and study the isolated effect of individual quality incentives. Effects over time will be studied to shed light on the mechanisms driving quality improvement.

Hygienic quality could be affected in various ways. The extra bonuses increased the quality premium β for hygienic quality that farmers are confronted with. Cost of effort may have decreased for at least three reasons: (*a*) farmers received physical capital in the form of buckets, filters, and milk can loans; (*b*) trainings provided farmers with new knowledge; and (*c*) trainings channeled scarce bandwidth (mental capacity and executive control) to the issue of hygienic milk production (Schilbach, Schofield, and Mullainathan 2016; Treurniet 2020). Finally, introducing an individual incentive system increased ω_i to 1. To jointly test the three hypotheses, I will study whether this bundle of interventions contributes to an increase in the hygienic quality of milk.

Last, although behavioral changes can quickly affect hygienic quality, intervention impacts might change over time. On the one hand, the physical and human capital inputs may decay over time. For example, filters might become dirty and buckets might be damaged. Although the hygiene training may temporarily channel bandwidth to the issue of hygienic milk production, the bandwidth channeled to the issue of hygienic milk production may become thinner again after completion of the training. This makes the quality improvement more costly. On the other hand, learning by doing might decrease the marginal cost of quality improvement. Whether impacts will decrease or increase over time is thus theoretically ambiguous. I will study which effect dominates by exploring whether impacts on hygienic quality change over time.

Data

The data used in this article are derived from three sources: (*a*) a baseline survey, (*b*) cooperative administrative data, and (*c*) additional quality tests conducted. A combination of the first two data sources was used to control for pre-existing differences, whereas a combination of the latter two was used to construct outcome variables.

[5]As solids increase during milking, throwing away first milk slightly increases compositional quality, but a large share of the milk should be thrown away to substantially increase compositional quality.

Baseline Variables

To control for socio-economic differences across farmers at upgraded and non-upgraded MCPs, variables were taken from a baseline survey. During October and November 2015 all farmers delivering to thirteen selected MCPs, including six of the seven MCPs that would be upgraded, were selected to be visited in their homes for a baseline survey with questions on general demographics, household assets, number of dairy cows, farm assets, labor, and trust.[6,7,8] Out of a sample list of 1,351 farmers, 1,335 farmers (98.8%) were surveyed. Respondents were asked for their member ID, so we could associate baseline survey data with records in cooperative administrative data. Out of the 1,335 farmers surveyed at baseline, thirty-nine farmers (2.9%) were reported to hold more than one ID. In these cases, I ignored extra IDs, as they were more likely to be held temporarily and to be transferred to other farmers during the study period.[9]

To control for the quantity and quality of milk delivered to the cooperative, several indicators were obtained from cooperative administrative data for 2014, the year preceding the start of the intervention, with records for every ten days and all 3,682 farmers that delivered to the cooperative.[10] The quantity of milk was measured at the individual level for each delivery, and I calculated the sum over the whole year to get the total milk delivered per farmer.[11] To obtain proxies for quality, samples of milk were taken twelve times per month. For budgetary reasons, samples were taken per payment group instead of per individual farmer. These group samples were subsequently tested in the cooperative's laboratory.

For compositional quality, two indicators were available. First, the total solids content measures the solid constituents fat, protein, and lactose. I used the mean of total solids measures over all periods in which the farmer delivered milk as primary indicator for compositional quality. Second, because cows typically produce milk with a freezing point around −0.540°C and adding water raises the freezing point, the freezing point of milk delivered can be used as an indicator for added water (Shipe 1959). Farmers received a small penalty on their milk price if the payment group's freezing point was above −0.520°C and a larger penalty if it was above −0.500°C. To limit the impact of outliers in the continuous freezing point variable, I used the frequency of the freezing point exceeding these thresholds as secondary proxy for compositional milk quality.

For hygienic quality, one proxy was available. The cooperative performed a Resazurin test, in which the color of milk after a controlled chemical process indicates the degree of bacterial contamination. Farmers received a bonus if the average Resazurin grade of their payment group was supposed to correspond to a TPC below the Indonesian National Standard of 1 million cfu/mL. I used the frequency of this bonus being applied as proxy for hygienic quality.

Outcome Variables

Total solids content and freezing points, as indicators for compositional quality, are again taken from cooperative administrative data and are available in panel format until December 2018.[12] Although these compositional quality indicators are measured at individual farmer level at upgraded MCPs, they are measured at payment group level at non-upgraded MCPs. I use the total solids content and a binary indicators specifying whether the freezing point exceeds −0.520°C.[13]

To proxy for hygienic milk quality, I use TPC results from individual sampling and more precise Bactoscan tests, which were introduced at upgraded MCPs and are standard in the global milk processing industry.

[6]The thirteen MCPs were selected for surveys by evaluators such that they included five MCPs that were pre-selected to be upgraded and eight other MCPs that scored similarly on crude MCP-level variables, like the number of farmers delivering and total milk quality delivered to these MCPs, while making sure that there was some geographical spread across the area in which the cooperative was active.

[7]The survey sample was selected at the level of administrative units. Most administrative units had their own MCP building, but some administrative groups shared a MCP building. At one of the MCP buildings, only one of two administrative units was selected.

[8]Details on the construction of survey index variables are included in online supplementary appendix table A1.

[9]In December 2017 and January 2018, slightly more than two years after the baseline survey, less than 1% of the main IDs was held by a different household, either temporarily or permanently.

[10]Data are missing for June 2014.

[11]To proxy total quantity delivered in 2014, I multiplied the total quantity delivered in the remaining months by the factor 12/11. I made a similar correction to the milk income variables.

[12]Until August 2016, record are aggregated per ten days. As of September 2016, record are aggregated per fifteen days. Data are missing for the first half of October 2016.

[13]I do not show results for the frequency of the freezing point being above −0.500°C, as this occurred very rarely at both upgraded and non-upgraded MCPs during the endline period.

To obtain similar TPC measures for farmers at non-upgraded MCPs, one individual sample was taken for every farmer that delivered to the non-upgraded MCPs in our baseline survey sample and also tested with the Bactoscan. These additional quality measures were collected between April and July 2018, so that they can be compared with quality measures from upgraded MCPs that were collected in the same period. I use the TPC as well as two transformations: the logarithm and a binary indicator specifying whether the TPC is below the Indonesian National Standard of 1 million cfu/mL.

After I made variable transformations, I calculated weighted means over multiple measures collected during the endline period, as I will discuss in more detail in the next section.

Weighting Variable

For farmers at non-upgraded MCPs, I only have measures for compositional quality from a group sample, which is a weighted average of individual quality. I will thus compare individual measures from farmers at upgraded MCPs, with weighted group averages from farmers at non-upgraded MCPs. If individual quality is correlated with milk quantity delivered, the ordinary mean deviates from the weighted mean, causing inference in non-weighted regressions to be biased. For the compositional quality, I therefore only show comparisons weighted by milk quantity delivered. For non-transformed variables, like the total solids content, the weighted sum of individual measures in expectation equals the group measure, so weighting by milk quantity delivered solves the issue. For transformed variables, like the binary freezing point indicator, the weighted sum of individual measures may be different from the group measure in general. As I will discuss in the results section, however, they are very similar in my case.

As I have individual measures for hygienic quality at endline, I am able to show both farmer-level comparisons (all farmer have equal weight) and milk-level comparisons (weighted by milk quantity delivered). Whereas in the main text a balance table is only shown for farmer-level comparisons, I also present impact regressions results of the regressions weighted by milk quantity for comparison with the previous section and because of relevance for policy. Farmer-level results do not structurally differ from milk-level results.

The quantity of milk produced is taken from cooperative administrative data and available in panel format until December 2018. The quantity is measured at individual level for all farmers.

Sample

The sample consists of all 1,335 cooperative members that participated in our baseline survey, excluding (*a*) twenty-four farmers that did not deliver at least some milk in the baseline year 2014, (*b*) four farmers that delivered their last milk in 2014 to another MCP than the thirteen selected for this study, and (*c*) one cooperative board member and sixteen participants in a small other training program, which could have benefitted from other sources of information. From the remaining 1,290 farmers, 226 farmers did not deliver at least some milk to the cooperative during our endline period February–December 2018. This attrition was not significantly associated to the intervention, and these farmers were left out of further analysis. This leaves a sample of 1,064 farmers.

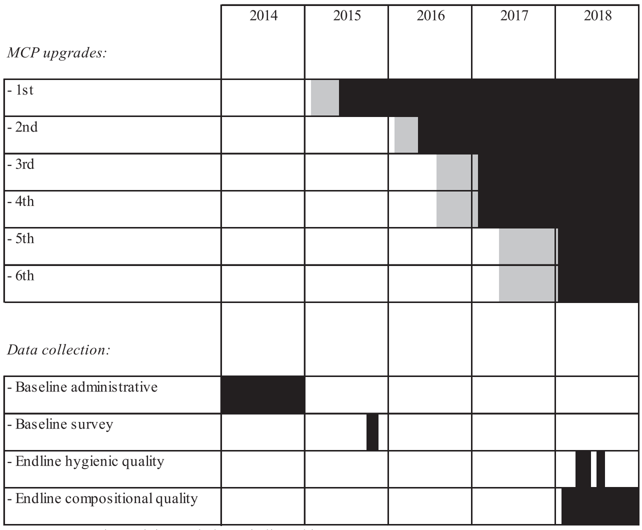
Timing

Table 1 shows the timing of data collections together with the implementation of the intervention at the upgraded MCPs in my sample. As can be seen, the baseline survey took place after the opening of the first upgraded MCP and might have been affected by the intervention. Further, the collection of endline quality indicators used in my impact regression analysis started quickly after the opening of the last two upgraded MCPs, thus limiting the study of longer term impacts. These concerns will be addressed in several robustness checks.

Empirical Strategy for Compositional Quality

Which MCPs are upgraded is determined by the public–private partnership and therefore endogenous. Upgraded MCPs are larger in terms of number of farmers and total quantity delivered, and farmers at upgraded MCPs were more likely to be female, were more wealthy, and owned more cows and farm assets. However, four factors help me to attribute causality. First, pre-intervention differences in quality were not among the criteria to select MCPs, and I do not find any

Table 1. Timing of Intervention and Data Collection



Note: Intervention training periods are indicated in gray.

significant difference in the pre-intervention quality variables. Second, the intervention is assigned at MCP level and not at individual level, so I am able to match farmers that delivered to upgraded MCPs to similar farmers that delivered to non-upgraded MCPs. Third, the availability of monthly panel data allows to study effects over time. As upgraded MCPs opened at different times, I am able to study whether increases in quality at these MCPs occur shortly after the opening of upgraded MCPs. Fourth, the data allow for various robustness checks.

In the previous section, I already discussed how I weight observations by milk quality delivered to correct for differences in the level of measurement. In this section, I discuss the empirical strategy for the isolated impact of individual quality incentives on compositional milk quality. I start by discussing my matching strategy and subsequently explain how I will construct graphs and estimate impact regression models. After discussing results for compositional quality in the next section, I introduce some robustness checks.

Matching

Online supplementary appendix table A2 includes baseline summary statistics of milk deliveries to MCPs that were later upgraded and milk deliveries to MCPs that would remain non-upgraded. At the bottom, the table also shows the distribution of available quality measures over months. Farmers at upgraded MCPs were more likely to be female, were more wealthy, had more cows, and had more farm assets, but there are no significant pre-existing differences in the quality of milk delivered by farmers at upgraded and non-upgraded MCPs. Quality measures were equally available for all months in the endline period.

To obtain appropriate comparison groups, I employed a Coarsened Exact Matching procedure as discussed by Iacus, King, and Porro (2012). This procedure reweights observations from non-upgraded MCPs to mimic the distribution of observations at upgraded MCPs for pre-selected matching variables. I exactly matched milk on (*a*) gender,[14] (*b*) presence in one of six equidistant intervals of the International Wealth Index, (*c*) number of cows, and (*d*) number of farm assets.

[14]Qualitative research by Wijers (2019) indicates that although women are generally important actors in smallholder milk production in Indonesia, their roles are often not formalized within the cooperative. As women who are formally responsible for dairy farming might come from households with a different intra-household work division, I control for gender in my analysis.

The exact matching procedure creates very similar matches, but due to the curse of dimensionality a substantial part of the sample remains unmatched. Given that I started with a large sample, a large sample remains, so statistical power does not seriously decrease. Further, because many of these unmatched respondents are not very different from matched ones, the curse of dimensionality has only a minor effect on external validity. However, the matching procedure may drop respondents, whose milk still counts in the group measure of others, and this issue will be addressed in a robustness check.

Because quality measures are equally available over time, there is no need to control for differential seasonal effects. Yet, for consistency with later analysis, I corrected for potential seasonal effects by calculating for each farmer a weighted average over available quality indicators, with weights for farmers at to-be-upgraded MCPs chosen such that the distribution over months for farmers at intervention MCPs equals the distribution for farmers at comparison MCPs.

Table 2 shows balance at baseline after matching. The intervention groups are now very similar. There is no difference in variables directly used in the matching procedure, whereas differences in other variables are limited.

Graphical Evidence

I have then graphed the evolution of quality over time for deliveries to upgraded MCPs and their matched deliveries to non-upgraded MCPs. Because MCPs are upgraded at different times, I show one graph for each upgraded MCP.[15]

Impact Regressions

To formally test the effect of individual incentives on compositional quality, I regressed the mean total solids (TS) content and the frequency of the freezing point (FP) being above −0.520°C on the intervention indicator. As control variables, I included the baseline characteristics that were listed in table 2, including baseline indicators for both compositional quality variables.

More formally, I estimated:

$$(5) \quad q_{ijc} = \alpha + \beta Upgraded_c(+ \gamma X_{ic}) + \varepsilon_{ic}$$

where q_{ic} is the quality outcome available for farmer i, which at baseline delivered to MCP c, calculated as the mean for the endline period (see table 1); $Upgraded_c$ equals 1 if MCP c is upgraded at endline and 0 otherwise; and X_{ic} is the vector of control variables, which is included in half of the regressions. Standard errors are clustered at the MCP level. Because the number of MCPs is small, bootstrapped p-values are obtained with the wild bootstrap using Rademacher weights and imposing the null hypothesis as proposed by Cameron, Gelbach, and Miller (2008).

Table 2 indicates that non-compliance by farmers is limited. Within the matched sample, both switching from intervention MCPs to comparison MCPs and vice versa was rare. To correct for these limited individual selection effects, I estimated Intention-To-Treat (ITT) effects by using as intervention indicator the actual endline upgrade status of the MCP to which the farmer delivered its last milk in 2014. As non-compliance to the treatment status of the this MCP is limited, the ITT effect is close to average causal impact of the intervention at upgraded MCPs (Average Treatment Effect on the Treated, ATT), where causal attribution relies on the assumption that matched famers at non-upgraded MCPs provide an accurate counterfactual for farmers at upgraded MCPs.

Results for Compositional Quality

Figures 2 and 3 show the evolution of compositional quality indicators over time for farmers at upgraded MCPs and their comparison groups, broken down by upgraded MCP. Though trends in total solids content differ across upgraded MCPs, a substantial and quick increase in the total solids content is observed at all upgraded MCPs around their opening. The quick responses around varying opening times increase confidence in causal attribution to the intervention. Over the years, the frequency that the freezing point exceeded −0.520°C decreased but only after the upgrade of MCPs, freezing points above −0.520°C virtually disappeared for a longer period of time.

Given the quick responses in compositional quality, the increase in compositional quality

[15] Table 2 indicated that a very small proportion of farmers switched from to-be-upgraded MCPs to not-to-be-upgraded MCPs. Although these farmers are included in the Intention-to-treat impact regressions, they are not shown in figures 2 and 3.

Table 2. Balance at Baseline after Matching (Comparing Milk across Intervention Status)

	Upgraded (ITT)			Non-upgraded (ITT)			Diff
	N	Mean	*SD*	N	Mean	*SD*	p
Age of respondent	381	44.10	12.01	374	44.15	10.18	0.978
Gender of respondent	381	0.072	0.259	374	0.072	0.259	1.000
Junior high school or higher	381	0.297	0.458	374	0.304	0.461	0.960
Household asset index	381	4.072	1.540	374	4.401	1.653	0.168
International Wealth Index	381	67.62	10.60	374	67.24	10.77	0.842
Progress out of Poverty Index	381	38.93	7.92	374	39.18	8.66	0.878
Number of dairy cattle total	381	4.862	2.243	374	4.862	2.243	1.000
Farm asset index	381	2.738	0.617	374	2.738	0.617	0.957
Number of non-family full-time workers	381	0.049	0.267	374	0.024	0.160	0.236
Number of non-family part-time workers	381	0.029	0.217	374	0.004	0.067	0.146
Number of family full-time workers	381	1.130	0.685	374	1.226	0.613	0.150
Number of family part-time workers	381	0.642	0.531	374	0.613	0.505	0.682
Milk quantity (1,000 L)	381	10.235	6.194	374	9.896	5.468	0.788
Milk quantity including extra IDs (1,000 L)	381	10.521	6.693	374	10.240	6.039	0.794
Milk compositional quality TS (%)	381	11.76	0.28	374	11.76	0.27	0.972
Milk compositional quality FP (> −0.520°C)	381	0.172	0.209	374	0.189	0.227	0.694
Milk compositional quality FP (> −0.500°C)	381	0.012	0.040	374	0.007	0.036	0.520
Milk hygienic quality TPC (bonus)	381	0.553	0.280	374	0.610	0.306	0.454
Milk price (1,000 IDR)	381	4.153	0.147	374	4.155	0.145	1.000
Milk income (1,000,000 IDR)	381	42.63	26.07	374	41.23	22.96	0.790
Milk income including extra IDs (1,000,000 IDR)	381	43.83	28.11	374	42.68	25.48	0.800
Payment group size	381	6.245	4.111	374	6.602	3.492	0.824
Distance to MCP (km)	381	0.858	0.820	374	0.784	0.829	0.812
Trust index[a]	370	3.619	0.367	371	3.667	0.345	0.314
Children want to take over	381	0.703	0.344	374	0.667	0.348	0.532
Upgraded at endline	381	0.993	0.086	374	0.010	0.100	0.000
Weight quality measures Feb-2018	381	0.091	0.019	374	0.091	0.023	0.992
Weight quality measures Mar-2018	381	0.091	0.016	374	0.091	0.019	0.976
Weight quality measures Apr-2018	381	0.091	0.013	374	0.091	0.020	1.000
Weight quality measures May-2018	381	0.091	0.009	374	0.091	0.017	0.980
Weight quality measures Jun-2018	381	0.091	0.009	374	0.091	0.016	0.980
Weight quality measures Jul-2018	381	0.091	0.011	374	0.091	0.017	0.973
Weight quality measures Aug-2018	381	0.091	0.011	374	0.091	0.016	1.000
Weight quality measures Sep-2018	381	0.091	0.010	374	0.091	0.016	1.000
Weight quality measures Oct-2018	381	0.090	0.012	374	0.090	0.046	1.000
Weight quality measures Nov-2018	381	0.091	0.012	374	0.091	0.017	0.998
Weight quality measures Dec-2018	381	0.090	0.014	374	0.090	0.047	0.990

Note: Bootstrapped p-values obtained with wild bootstrap using Rademacher weights and imposing the null hypothesis as proposed by Cameron, Gelbach, and Miller (2008).
[a]For later regression analysis, missings are set to the mean of farmers with the same upgrade status (ITT).

is most likely explained by decreasing added water. As impacts are not observed to increase over time, improved feed and cow health are not likely to have substantially contributed to better compositional quality. The impacts found on compositional quality are thus unlikely to be a byproduct of the intervention's efforts to increase the hygienic quality of milk but instead the result of introducing individual incentives for compositional quality.

Table 3 shows impact regression results. In line with hypothesis 1, both indicators reveal a significant improvement. Estimates suggest that the total solids content increased on average by 3.5–3.8% relative to the counterfactual mean, which corresponds to 1.39–1.54 standard deviations. Because processors generally do not value water content, this suggests that the value of milk increased by 3.5–3.8%. The frequency of the freezing point being above

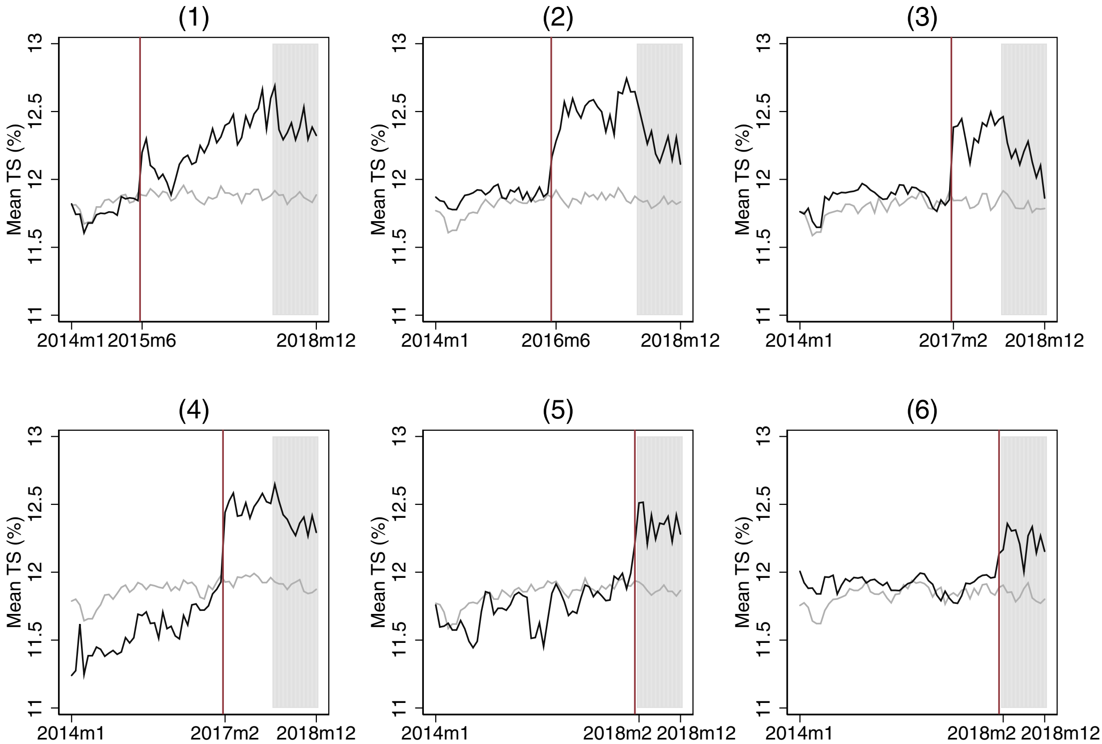


Figure 2. Evolution of total solids (TS) content of farmers at upgraded MCPs (black line) and matched farmers at non-upgraded MCPs (gray line) over time.

Note: Gray shaded months indicate measures used in the impact regressions. Vertical lines indicate the opening of the upgraded MCP

–0.520°C almost decreased to zero, which confirms that the water content was lower.

Robustness Checks

As discussed earlier, the weighted sum of transformed individual measures may be different from the transformation of the group measure in general. However, in this study we see that the frequency that the freezing point exceeds –0.520°C is virtually zero at upgraded MCPs. If all members of a hypothetical group deliver milk with a freezing point below –0.520°C, then the group's combined milk will also have a freezing point below –0.520°C. For hypothetical groups at upgraded MCPs, the frequency that the freezing point exceeds –0.520°C would thus also equal almost zero, so results would be very similar if I would have had group-level measures from upgraded MCPs similar to those from non-upgraded MCPs.

Although the selection of matching variables directly follows from the balance table, one might wonder whether the regression results are caused by my specific matching procedure, which could increase concerns for p-hacking (Brodeur, Cook, and Heyes 2020). As a robustness check, I therefore re-estimated the main results in table 3 while skipping the matching stage. Results are included in online supplementary appendix table A3 and are similar to the main results.

As an additional robustness check on the matching procedure, I have used the synthetic control method to construct counterfactual estimates that mimic the evolution of total solids over time (Abadie and Gardeazabal 2003; Abadie, Diamond, and Hainmueller 2010). I again started with my sample of 1064 farmers for which I have observations in both the baseline and endline period. As individual quality measures include quite some transitory shocks and I want to avoid constructing more or less random matches, I have first aggregated observations by MCP by calculating means over all farmers weighted by quantity delivered. Subsequently, I have constructed one synthetic control for each

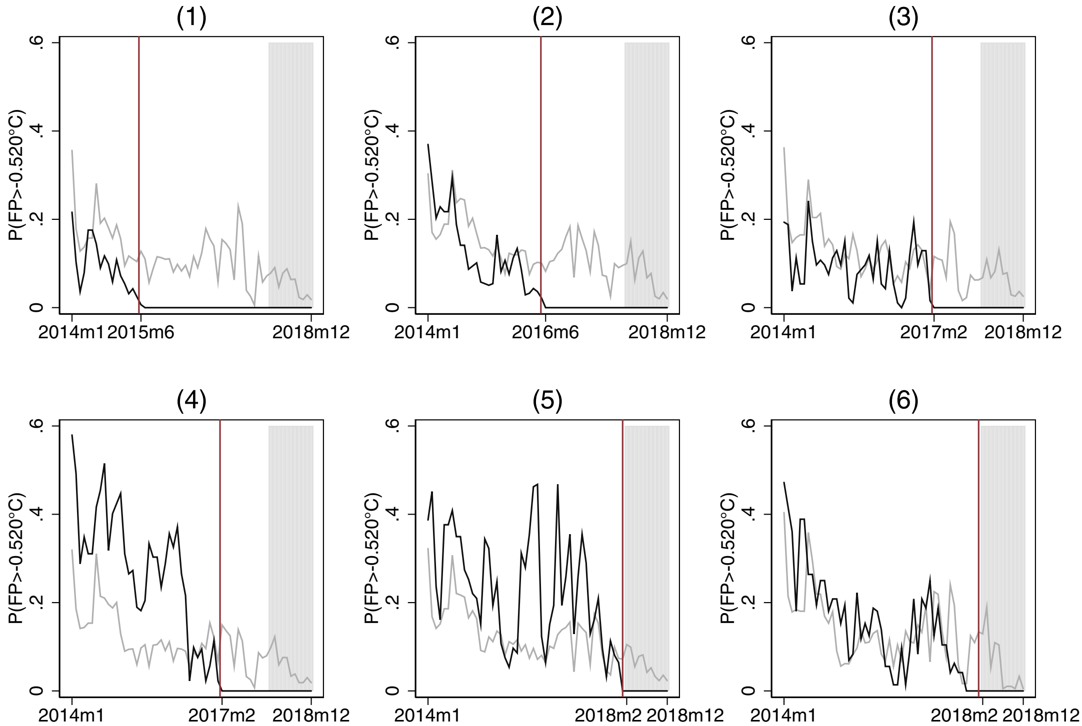


Figure 3. Evolution of the frequency of the freezing points (FP) exceeding −0.520°C of farmers at upgraded MCPs (black line) and matched farmers at non-upgraded MCPs (gray line) over time.

Note: Gray shaded months indicate measures used in the impact regressions. Vertical lines indicate the opening of the upgraded MCP

upgraded MCP as a weighted average of the seven non-upgraded MCPs in my sample. The weights were chosen to minimize the sum of squared differences in total solids between the upgraded MCP and the synthetic control over all months before the training started at the upgraded MCP. Online supplementary appendix figure A1 reports results and shows a positive effect for all six upgraded MCPs in my sample. When I repeat this analysis with all farmers in the monitoring data that delivered milk in both the baseline and

Table 3. Impact on Compositional Quality

	Outcome variables			
	(1)	(2)	(3)	(4)
	TS (%)	TS (%)	FP > −0.520°C	FP > −0.520°C
Upgraded (ITT)	0.410***	0.453***	−0.0344***	−0.0348***
Clustered SE	(0.080)	(0.070)	(0.0103)	(0.0101)
Bootstrapped p-value	0.000	0.000	0.010	0.004
Baseline controls	No	Yes	No	Yes
Level of comparison	Milk	Milk	Milk	Milk
MCPs	13	13	13	13
Observations	755	755	755	755
Mean of non-upgraded	11.870	11.870	0.0346	0.0346
SD of non-upgraded	0.295	0.295	0.0837	0.0837

Note: Table shows comparison after weighting observations by milk quantity. Standard errors clustered at MCP level in parentheses. Bootstrapped p-values obtained with wild bootstrap using Rademacher weights and imposing the null hypothesis as proposed by Cameron, Gelbach, and Miller (2008). Asterisks indicate the following:
***p < 0.01.

endline period, including those that were not surveyed,[16] I find similar effects (see online supplementary appendix figure A2). Using this larger sample further allows me to run placebo tests with twenty-two non-upgraded MCPs, in which the non-upgraded MCPs are iteratively assigned to treatment and the upgraded MCP joins the remaining non-upgraded MCPs. The resulting differences with the synthetic control are shown in online supplementary appendix figure A3, and are clearly outstanding for upgraded MCPs. Following Abadie, Diamond, and Hainmueller (2010), I formally calculate the ratio of the MSPE after the opening of the upgraded MCP and the MSPE before the training started at this MCP and compare the ratio for upgraded MCPs with those of the distribution of ratios generated by placebos. For each of the first five upgraded MCPs, this ratio is larger than for any of the placebo MCPs, which leads to a p-value of 1/23 = 0.043. For the sixth upgraded MCP, the p-value equals 3/23 = 0.130. These results confirm my earlier finding that the introduction of individual incentives increased the total solids of milk delivered by farmers.

As the baseline survey was held after the opening of the first upgraded MCP, one might fear that some baseline survey characteristics are affected by the intervention, thus invalidating the construction of comparable intervention groups. As a robustness check, I therefore repeated my empirical procedure while excluding the first upgraded MCP.[17] Results are included in online supplementary appendix table A4 and are again similar to the main results.

To study whether impacts are similar for the MCPs that are open for at least a year, I repeated my empirical procedure while excluding the last two upgraded MCPs, which opened in 2018. Results are included in online supplementary appendix table A5 and are very similar to the main results.

Finally, one might fear that missing group members and matching invalidate inference on group-level quality measures. To address this concern, I excluded incomplete groups, which had one or more members missing from the matched sample, either because (*a*) they did not deliver milk in the baseline year 2014, (*b*) they did not participate in the baseline survey or (*c*) no good match was found. Instead of matching farmers at non-upgraded MCPs to farmers at upgraded MCPs, I matched farmers at upgraded MCPs to farmers at non-upgraded MCPs, so that the sampling weights for farmers from non-upgraded MCPs are not interacted with matching weights. Results are included in online supplementary appendix table A6 and are again similar.

Empirical Strategy for Hygienic Milk Quality

I now turn to the empirical strategy for the impact on hygienic quality. As discussed, I study the combined effect of the physical upgrade, intervention trainings, and individual quality incentives, and how this effect changes over time. The structure mimics the structure of the previous section, and the discussion focuses on deviations from the strategy employed before.

Matching

Online supplementary appendix table A7 includes baseline summary statistics of milk deliveries to MCPs that were later upgraded and milk deliveries at MCPs that would remain non-upgraded. Although the samples slightly differ due to data availability, the sample is quite similar as before, causing the same patterns to be observed. Farmers at non-upgraded MCPs were slightly more likely to have their quality being tested in later months, although differences are not statistically significant.

To obtain appropriate comparison groups, I employed the same coarsened exact matching procedure with the same matching variables as before. I also used the same strategy to correct for potential seasonal effects.

Table 4 shows balance at baseline after matching. The intervention groups are again very similar.

Impact Regressions

To estimate the effect of the intervention on hygienic quality, I regressed the TPC, the natural logarithm of TPC, and the frequency of the TPC being below the Indonesian National

[16] Out of the population of twenty-four non-upgraded MCPs, I exclude one MCP where the cooperative implemented similar policy changes and one MCP from which almost all farmers switched to upgraded MCPs.

[17] In this and some other robustness checks, after matching, the number of non-family full-time workers was found to be significantly larger for farmers from upgraded MCPs, whereas the absolute difference was still small. If anything, principal-agency challenges might make it harder to increase quality for farmers with non-family workers.

Table 4. Balance at Baseline after Matching (Comparing Farmers across Intervention Status)

	Upgraded (ITT)			Non-upgraded (ITT)			Diff
	N	Mean	*SD*	N	Mean	*SD*	p
Age of respondent	356	44.81	12.49	343	44.79	10.60	0.930
Gender of respondent	356	0.070	0.256	343	0.070	0.256	0.984
Junior high school or higher	356	0.267	0.443	343	0.301	0.460	0.724
Household asset index	356	3.944	1.485	343	4.274	1.598	0.144
International Wealth Index	356	66.42	10.86	343	66.04	10.80	0.760
Progress out of Poverty Index	356	38.51	8.28	343	38.51	8.80	0.994
Number of dairy cattle total	356	4.430	2.138	343	4.430	2.138	1.000
Farm asset index	356	2.697	0.622	343	2.697	0.622	0.974
Number of non-family full-time workers	356	0.053	0.271	343	0.022	0.171	0.106
Number of non-family part-time workers	356	0.011	0.130	343	0.004	0.060	0.688
Number of family full-time workers	356	1.135	0.662	343	1.204	0.585	0.308
Number of family part-time workers	356	0.626	0.534	343	0.592	0.511	0.676
Milk quantity (1,000 L)	356	8.715	5.558	343	9.014	5.188	0.750
Milk quantity including extra IDs (1,000 L)	356	8.894	5.848	343	9.288	5.722	0.676
Milk compositional quality TS (%)	356	11.72	0.29	343	11.74	0.27	0.868
Milk compositional quality FP (> −0.520°C)	356	0.212	0.237	343	0.215	0.241	0.964
Milk compositional quality FP (> −0.500°C)	356	0.018	0.052	343	0.008	0.033	0.326
Milk hygienic quality TPC (bonus)	356	0.511	0.287	343	0.576	0.308	0.416
Milk price (1,000 IDR)	356	4.127	0.156	343	4.141	0.147	0.754
Milk income (1,000,000 IDR)	356	36.10	23.20	343	37.44	21.78	0.740
Milk income including extra IDs (1,000,000 IDR)	356	36.85	24.42	343	38.59	24.13	0.648
Payment group size	356	6.204	3.970	343	6.800	3.500	0.752
Distance to MCP (km)	356	0.902	0.835	343	0.789	0.810	0.674
Trust index[a]	346	3.605	0.375	340	3.648	0.356	0.324
Children want to take over	356	0.691	0.356	343	0.656	0.360	0.456
Upgraded at endline	356	1.000	0.000	343	0.014	0.116	0.000
Weight quality measures Apr-2018	356	0.260	0.253	343	0.260	0.437	1.000
Weight quality measures May-2018	356	0.300	0.295	343	0.300	0.457	1.000
Weight quality measures Jul-2018	356	0.439	0.305	343	0.439	0.494	1.000

Note: Bootstrapped p-values obtained with wild bootstrap using Rademacher weights and imposing the null hypothesis as proposed by Cameron, Gelbach, and Miller (2008).

[a]For later regression analysis, missings are set to the mean of farmers with the same upgrade status (ITT).

Standard of 1 M cfu/mL on the intervention indicator. I included the same control variables as before, which include the baseline indicator for TPC. Because knowledge was likely to spill over to non-upgraded MCPs via regular activities of cooperative extension officers, my results underestimate effects of the knowledge component of the intervention and are therefore a conservative estimate of the total intervention impact.

Graphical Evidence

Because I have only one TPC measure per farmer at non-upgraded MCPs, the study of evolution of hygienic quality over time is limited to upgraded MCPs.

Results for Hygienic Quality

Table 5 shows impact regression results. The hygienic quality significantly improved, as the TPC decreased by about one third, and the average TPC at upgraded MCPs now satisfies the national standard (columns 1–2). Because bacteria grow exponentially, the effect may be more precisely estimated when I first take the logarithm of TPC as dependent variable. When doing this, the results are similar (columns 3–4). The TPC of individual deliveries was also significantly more likely to be below the national standard (columns 5–6). Milk-level results are similar to the results of the farmer-level comparison.

Table 5. Impact on Hygienic Quality

	Outcome variables					
	(1) TPC	(2) TPC	(3) Log(TPC)	(4) Log(TPC)	(5) TPC < 1M	(6) TPC < 1M
Upgraded (ITT)	−414,318.9**	−404,451.1**	−0.317***	−0.294***	0.139***	0.135***
Clustered SE	(133,605.0)	(106,018.1)	(0.098)	(0.054)	(0.040)	(0.030)
Bootstrapped p-value	0.016	0.014	0.008	0.004	0.000	0.004
Baseline controls	No	Yes	No	Yes	No	Yes
Level of comparison	Farmer	Farmer	Farmer	Farmer	Farmer	Farmer
MCPs	13	13	13	13	13	13
Observations	699	699	699	699	699	699
Mean of non-upgraded	1,177,466.5	1,177,466.5	13.574	13.574	0.708	0.708
SD of non-upgraded	1,516,705.6	1,516,705.6	0.908	0.908	0.455	0.455
Upgraded (ITT)	−420,193.9**	−384,799.9***	−0.320***	−0.294***	0.135**	0.125***
Clustered SE	(132,503.6)	(66,827.2)	(0.103)	(0.045)	(0.043)	(0.030)
Bootstrapped p-value	0.014	0.002	0.008	0.006	0.010	0.000
Baseline controls	No	Yes	No	Yes	No	Yes
Level of comparison	Milk	Milk	Milk	Milk	Milk	Milk
MCPs	13	13	13	13	13	13
Observations	699	699	699	699	699	699
Mean of non-upgraded	1,126,619.7	1,126,619.7	13.556	13.556	0.725	0.725
SD of non-upgraded	1,461,142.4	1,461,142.4	0.883	0.883	0.446	0.446

Note: Top panel shows comparison at farmer level, bottom panel shows comparison after weighting observations by milk quantity. Standard errors clustered at MCP level in parentheses. Bootstrapped p-values obtained with wild bootstrap using Rademacher weights and imposing the null hypothesis as proposed by Cameron, Gelbach, and Miller (2008). Asterisks indicate the following:
**$p < 0.05$.
***$p < 0.01$.

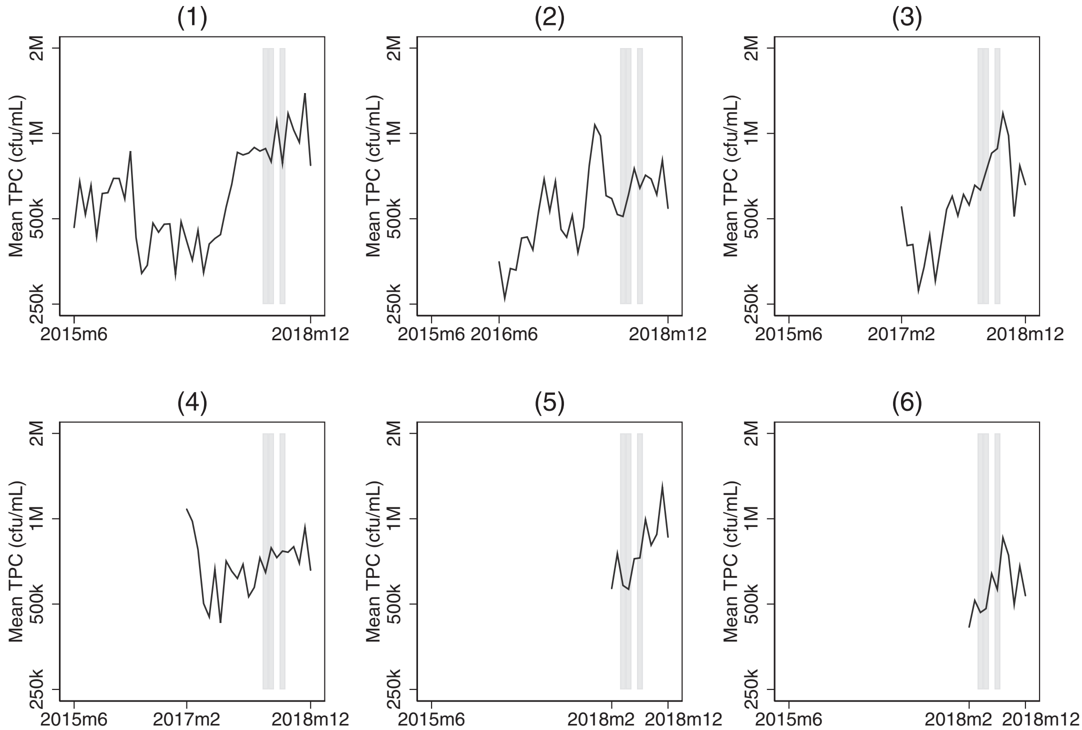


Figure 4. Evolution of TPC of farmers at upgraded MCPs over time.

Note: Gray shaded months indicate measures used in the impact regressions

Figure 4 shows the evolution of hygienic quality for all farmers in the matched sample per upgraded MCP.[18] The hygienic quality decreased and the TPC increased over time after the opening of the upgraded MCP, suggesting that the impact of the intervention on hygienic quality decreased over time. As this was seen as a problem by the processor and the cooperative extension officers again visited all farmers with checklists. They frequently found that farmers used dirty filters or inappropriate buckets. Addressing these problems would be easy and cheap but had not received sufficient attention of farmers. As the period of endline quality indicators used in my impact regression analysis starts soon after the opening of the last two upgraded MCPs, one might wonder whether impacts remain significantly when those two MCPs would be excluded from the analysis. This issue will be addressed in a robustness check.

Robustness Checks

Online supplementary appendix table A8 shows that qualitatively similar results are found when skipping the matching stage, suggesting that impact results do not critically depend on the matching procedure.

If farmers from the first upgraded MCP are excluded, results are still significant and point estimates are somewhat larger (see online supplementary appendix table A9). If farmers from the MCPs that were upgraded in 2018 are excluded, results are still significant and point estimates are somewhat smaller (see online supplementary appendix table A10).[19] Both results are in line with the suggestion that the impact of the intervention on hygienic quality decreased over time.

[18] Breakdown by MCP based on the upgraded MCP to which the farmer delivered first. I corrected for sample selection effects caused by missing values by replacing missing values by the most recent available measure (or the first available measure if no earlier measure was available).

[19] In the milk-level impact regression of the binary TPC measure without controls included, the intervention indicator is significant only at the 10% level.

Discussion

As the application of quality standards is increasing, smallholder farmers may not be able to meet standards and risk being excluded from modern value chains. New and cheaper opportunities to measure quality have the potential to reform small-scale transactions and might help farmers to keep up with increased quality standards.

This article finds that the introduction of an individual quality incentive system at a local dairy cooperative in Indonesia increased the compositional quality of milk. Together with physical inputs and training, individual quality incentives also increased the hygienic quality of milk. Results thus confirm that price incentives are a potent tool for quality improvement. This finding is in line with previous research (Saenger et al. 2013; Bernard et al. 2017; Hoffmann and Jones 2018; Magnan et al. 2019; Hoffmann et al. 2020).

The empirical setting in this article, however, is unique in that quality was not measured by a research body or a third party. Instead, quality was measured by the buyer, who could potentially save on the price that it pays by underreporting quality. For incentives to work well, however, smallholders need to trust that increasing quality will actually lead to better test results. The results in this article suggest that even if quality is measured by the buyer, price incentives have the potential to increase quality delivered by smallholder farmers, at least in the context of a dairy where farmers tend to be involved in relational contracts and receive signals on quality measurements on a regular basis.

From the moment that the intervention studied in this article started as a subsidized pilot, the individual incentive system has been maintained by the dairy processor and local cooperative already for years. Two factors may have contributed to this sustainability. First, once investments to implement the system were made, maintaining the system seemed to be in the economic interest of all value chain actors involved, and the value chain did not involve traders that would benefit from non-transparent transactions as was the case in the study of Bernard et al. (2017). Second, although in the past the cooperative had canceled penalties for high levels of TPC as members protested against the low prices that resulted for low quality milk, the intervention meetings stressed that farmers could use the intervention to earn higher prices. Moreover, as quality increased at upgraded MCPs, and the price function was not differentiated between upgraded and non-upgraded MCPs, farmers at upgraded MCPs earned higher prices than farmers at non-upgraded MCPs. This may have led cooperative members to accept potential low prices that can result from the individual incentive system.

Supplementary Material

Supplementary material are available at American Journal of Agricultural Economics *online.*

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association* 105: 493–505.

Abadie, Alberto, and Javier Gardeazabal. 2003. The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review* 93: 113–32.

Abate, Gashaw Tadesse, Banawe Plambou Anissa, Tanguy Bernard, Erwin Bulte, Alain de Janvry, and Elisabeth Sadoulet. 2019. Independent Quality Assessment and Certification of Wheat in Ethiopia: Smallholder Willingness to Pay, Quality and Financial Viability. Working Paper, Social Sciences Group, Wageningen University.

Abate, Gashaw Tadesse, and Tanguy Bernard. 2017. Farmers' Quality Assessment of Their Crops and Its Impact on Commercialization Behavior. IFPRI Discussion Paper 01624, Washington, DC.

Bernard, Tanguy, Alain de Janvry, Samba Mbaye, and Elisabeth Sadoulet. 2017. Expected Product Market Reforms and Technology Adoption by Senegalese Onion Producers. *American Journal of Agricultural Economics* 99: 1096–115.

Bernard, Tanguy, Melissa Hidrobo, Agnes Le Port, and Rahul Rawat. 2019. Nutrition-Based Incentives in Dairy Contract Farming in Northern Senegal. *American*

Journal of Agricultural Economics 101: 404–35.

Brodeur, Abel, Nikolai Cook, and Anthony Heyes. 2020. Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics. *American Economic Review* 110: 3634–60.

Burchardi, Konrad B, Selim Gulesci, Benedetta Lerva, and Munshi Sulaiman. 2019. Moral Hazard: Experimental Evidence from Tenancy Contracts. *Quarterly Journal of Economics* 134: 281–347.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller. 2008. Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics* 90: 414–27.

Casaburi, Lorenzo, and Rocco Macchiavello. 2015. Loyalty, Exit, and Enforcement: Evidence from a Kenya Dairy Cooperative. *American Economic Review* 105: 286–90.

Foster, Andrew D, and Mark R Rosenzweig. 2010. Microeconomics of Technology Adoption. *Annual Review of Economics* 2: 395–424.

Hoffmann, Vivian, and Kelly M Jones. 2018. Improving Food Safety on the Farm: Experimental Evidence from Kenya on Agricultural Incentives and Subsidies as Public Health Investments. IFPRI Discussion Paper 01746, Washington, DC.

Hoffmann, Vivian, Sarah Kariuki, Janneke Pieters, and Mark Treurniet. 2020. Safe Food for Me—and Maybe for You: Upside Risk, Premium Market Access, and Producer Demand for a Food Safety Technology. Working Paper, Social Sciences Group, Wageningen University.

Hoffmann, Vivian, Christine Moser, and Alexander Saak. 2019. Food Safety in Low and Middle-Income Countries: The Evidence through an Economic Lens. *World Development* 123: 104611.

Iacus, Stefano M, Gary King, and Giuseppe Porro. 2012. Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis* 20: 1–24.

Jack, B Kelsey. 2013. *Market Inefficiencies and the Adoption of Agricultural Technologies in Developing Countries*. White paper prepared for the Agricultural Technology Adoption Initiative, J-PAL (MIT)/CEGA (University of California at Berkeley). Available at: https://escholarship.org/uc/item/6m25r19c

Magnan, Nicholas, Vivian Hoffmann, Gissele Garrido, Faniel Akwasi Kanyam, and Nelson Opoku. 2019. Information, Technology, and Market Rewards: Incentivizing Aflatoxin Control in Ghana. IFPRI Discussion Paper 01878, Washington, DC.

Saenger, Christoph, Matin Qaim, Maximo Torero, and Angelino Viceisza. 2013. Contract Farming and Smallholder Incentives to Produce High Quality: Experimental Evidence from the Vietnamese Dairy Sector. *Agricultural Economics* 44: 297–308.

Saenger, Christoph, Maximo Torero, and Matin Qaim. 2014. Impact of Third-Party Contract Enforcement in Agricultural Markets—a Field Experiment in Vietnam. *American Journal of Agricultural Economics* 96: 1220–38.

Schilbach, Frank, Heather Schofield, and Sendhil Mullainathan. 2016. The Psychological Lives of the Poor. *American Economic Review* 106: 435–40.

Shipe, WF. 1959. The Freezing Point of Milk. A Review. *Journal of Dairy Science* 42: 1745–62.

Treurniet, Mark. 2020. The Impact of Being Surveyed on the Adoption of Agricultural Technology. Working Paper, Social Sciences Group, Wageningen University.

Wijers, Gea DM. 2019. Inequality Regimes in Indonesian Dairy Cooperatives: Understanding Institutional Barriers to Gender Equality. *Agriculture and Human Values* 36: 167–81.