IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18328

Quality Upgrading in the Street Food Market: Is Better Equipment and Training Sufficient?

Caitlin Brown
Denni Tommasi

DECEMBER 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18328

Quality Upgrading in the Street Food Market: Is Better Equipment and Training Sufficient?

Caitlin Brown
Université Laval

Denni Tommasi
University of Bologna, J-PAL, IZA and RFBerlin

DECEMBER 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Quality Upgrading in the Street Food Market: Is Better Equipment and Training Sufficient?*

We study quality upgrading in informal markets through two experiments with street-food vendors and consumers in India. First, we define quality in terms of *food safety* and develop a context-specific measurement framework. Second, we show that consumers are willing to pay substantial premiums for cleanliness. Third, we implement a vendor-level intervention that lowers upgrading costs and enhances the ability to signal quality through sanitation-related equipment. The intervention improves food-safety practices and profits, but effects are modest and fade over time. Fixed pricing norms and local environmental constraints appear central, consistent with a moral hazard model where cleanliness is not profitable.

JEL Classification: D82, I18, L15, L31, O12, O33

Keywords: quality upgrading, street food, informal markets, food safety, randomized experiment, consumer preferences, hygiene practices, moral hazard, subsidy effectiveness, signaling, developing countries

Corresponding author:
Denni Tommasi
University of Bologna
Piazza Scaravilli 2
Bologna, 40126
Italy

E-mail: denni.tommasi@unibo.it

* This project was funded by the United Nations Food and Agricultural Organization (FAO). We are extremely grateful to Cornelia Boesch, Markus Lipp, and Diletta Topazio who have supported us throughout the implementation, as well as their team at FAO's Food Safety Unit for several insightful conversations that helped shape the design and interpretation of the intervention. We also thank the program teams of Rajabazar Education and Awareness Development (READS), particularly the Secretary, Sarfaraz Ali, the Project Manager, Sudipta Moitra, the Field Supervisor, Samaresh Mukherjee, and the entire field team, for their instrumental role in facilitating the implementation of the project. Data collection was conducted by NYAS, and we are grateful to the entire team, with special thanks to Yashashvi Singh for his dedication. We are also indebted to InnoAid for their invaluable support in conducting this research project, in particular to Marie Louise Pollman-Larsen, Matias Pollmann-Larsen and Carmine Alaia. We acknowledge outstanding research assistance from Maria Battistin and Louis Steyn. The experiment has been registered with the AEA RCT registry (AEARCTR-0008797). IRB approval was obtained from the University of Bologna. Furthermore, we wish to express our gratitude to Kate Ambler, Andrew Clausen, Jishnu Das, Joshua Deutschmann, Selene Ghisolfi, Xavier Giné, Pawel Gola, Vivian Hoffmann, Eric Hsu, Erin Kelly, Gregory Lane, Annemie Maertens, David McKenzie, Dimitri Migrow, Sulagna Mookerjee, Mike Murphy, Sangyoon Park for insightful comments as well as seminar participants at BI Norwegian Business School, George Washington University, Georgetown University, HEC Montreal, Indian Statistical Institute (Kolkata), Inter-American Development Bank, Northeastern University, St Andrews University, Université libre de Bruxelles, University of Amsterdam, University of Bologna, Univer- sity of Edinburgh, University of Gothenburg, University of Lisbon, University of Pittsburgh, and the World Bank (both DECRG and DIME) for their valuable comments. We have no relevant or material financial interests related to the research described in this paper. All errors are our own.

1 Introduction

The persistence of low-quality goods in local markets remains a pervasive challenge in many developing countries. This issue is particularly concerning, as higher product quality can mitigate health risks, strengthen consumer confidence in local businesses, and serve as a catalyst for broader economic growth (Verhoogen, 2023). Much of the economic literature on this topic has focused on agricultural products, fertilizers, medicines, or services. Demand-side strategies—such as offering price premiums for high-quality products (Hoffmann and Jones, 2021; Hoffmann et al., 2023), fostering buyer–supplier partnerships (Park et al., 2023), providing credible information to consumers (Andrabi et al., 2017; Hasanain et al., 2023; Lane et al., 2023; Hsu and Wambugu, 2024), or enabling costly signaling mechanisms (Bai, 2024)—frequently succeed in enhancing product quality. In some markets, supply-side innovations, like the entry of a large, reputable firm, can independently shift the equilibrium toward higher quality (Bennett and Yin, 2019; Björkman Nyqvist et al., 2021). Conversely, smaller supply-side interventions, such as subsidies for safer inputs or new technology, tend to be effective only when paired with complementary demand-side measures (Bold et al., 2022; Deutschmann et al., 2023). Together, these findings demonstrate that tailored, context-specific interventions are essential for addressing the diverse market failures that allow low-quality goods to persist.

The interventions discussed above generally assume that, at least within experimental settings, it is possible to design incentives to *ensure* high-quality goods are credibly observed and rewarded before purchase. However, for certain types of goods deeply embedded in daily life, like food consumed outside the home, this assumption often *breaks down*. Transactions in these settings are frequent, low-value, and consumption typically occurs on the spot, making individual quality verification both impractical and costly. Moreover, these markets are highly fragmented, with countless small-scale producers operating independently, often with low levels of formal oversight and without links to more stringently regulated international supply chains. One sector that exemplifies this challenge is the informal street food sector.[1] Street food is a typical example of a credence good, where consumers cannot verify quality at the point of purchase—understood here primarily in terms of *food safety*—and may struggle to detect contamination even after falling ill. In such settings, food safety cannot be guaranteed in absolute terms or easily signaled to consumers. As a result, the goal of quality upgrading must shift from achieving consistent high standards to reducing the most salient contamination risks (FAO, 2017). This shift in focus raises a fundamental question: how can we design effective interventions in environments where quality is difficult to observe, incentives are weak, and enforcement is minimal? This challenge is especially pronounced in the street food sector, where documented health risks make quality failures particularly consequential (World Bank, 2019).[2]

Our analysis centers on two experiments conducted with street food consumers and vendors in Kolkata, India. We begin by collecting detailed baseline data on kiosk infrastructure, food safety inputs, and vendor practices.

[1] Street vending, the activity of selling goods and services in the streets without having a permanent built-up structure, is a large and growing sector across the developing world (Wongtada, 2014). Vendors comprise a substantial portion of urban informal employment, ranging from 2% to 24% in African, Asian, and Latin American cities (ILO, 2018). Among street vendors, food sellers represent the most visible group as they provide affordable food to 2.5 billion consumers every day, including up to 50% of daily energy and protein intake for those earning low- and middle-incomes (FAO, 2007).

[2] The WHO estimates that 600 million people suffer from foodborne diseases annually, resulting in 420,000 deaths and a loss of 33 million disability-adjusted life years (DALYs) due to contaminated food (WHO, 2022). Low-income countries in Asia and Africa account for 53% of all foodborne illnesses and 75% of related deaths. Precise global numbers on the proportion of foodborne disease from street food are hard to pin down, but available statistics, outbreak investigations, and microbiological studies give a sense of scale: various sources suggest that anywhere from roughly 5% up to 70% of foodborne illness cases in a community may stem from meals outside the home (including street food), depending on the region and local practices (e.g., Muinde and Kuria, 2005; Chukuezi, 2010; Muyanja et al., 2011; Samapundo et al., 2015; Abrahale et al., 2019; Gargiulo et al., 2022; Andrade et al., 2023).

The data confirm several key points consistent with previous assessments of this sector (e.g., Daniele et al., 2021). First, while many kiosks possess basic equipment needed for food preparation, these items are often in poor condition. Moreover, kiosks are typically located near busy roads, posing significant challenges to maintaining a location appropriate for the effective control of contaminants. Second, water, which is a critical input in kiosk operations, is generally of poor quality. Most vendors in our sample rely on public taps, and water samples from these sources show high contamination: 45% contain E. coli and 54% exceed recommended bacterial counts. Since very few vendors chlorinate their water, this raises serious health concerns. Third, compliance with essential practices of personal hygiene (and awareness more broadly) is also low: only 40% of vendors were observed washing hands with soap before handling food, and just 20% fully covered raw ingredients. Finally, we also collect detailed data on all items sold by vendors and find that prices are not systematically associated with vendors' equipment or hygiene practices.

In the first experiment, we examine the demand side to assess whether and to what extent consumers are willing to pay more for food prepared in more hygienic kiosks when presented with such a choice. This question is crucial because, without consumer demand for higher-quality goods, vendors have little incentive to invest in better equipment or adopt improved food safety practices. We survey 2,684 street food consumers who frequent vendors targeted for our supply-side intervention. Consumers often struggle to detect contaminated food, with only 41% reporting "it is easy to identify." Consumer food safety knowledge is limited, with respondents correctly identifying only a small number of common contaminants. On the other hand, 74% of respondents deem vendors' hygiene crucial when choosing a meal, and over half report that they would switch vendors after falling ill. To quantify preferences, we implement a Discrete Choice Experiment (DCE) where consumers choose between meals prepared at kiosks shown in photographs that differ in visual cleanliness and vendor hygiene. Although this is a stated preference approach, it is well suited to our context, where consumers must rely almost entirely on the visible appearance of kiosks and vendors to assess contamination risks at the time of purchase. Our results show that consumers are willing to pay significantly more for a meal from a cleaner-looking kiosk and from a vendor who appears hygienic. This highlights the importance for consumers of visible improvements in signaling food safety.

While encouraging, it remains unclear whether consumers would actually change their purchasing behavior when faced with more hygienic-looking options in real-world conditions. Translating this insight into effective action, however, is far from straightforward. Vendors face multiple barriers to maintaining consistently high food safety standards, two of which stand out. First, complying with general principles of food hygiene entails financial costs that small-scale vendors often cannot sustain. Second, many of the most important hygiene improvements are typically not visible to consumers, reducing the incentive to maintain them in the absence of external enforcement. These economic and informational frictions are further exacerbated by limited food safety knowledge—both in terms of understanding proper procedures and effectively communicating them to customers.

To address these two barriers, we design a second experiment involving 274 vendors that *reduces* the financial burden of higher-quality production while simultaneously *signaling* improvements to consumers through visible equipment upgrades. We geocode every kiosk and combine random assignment with proximity measures to distinguish own effects from spillovers to nearby vendors. Our approach leans more heavily on the supply side than previous studies by subsidizing many of the costs associated with producing higher-quality goods. This strategy compensates for the limited scope of demand-side interventions in this context, where consumers can observe visible infrastructure improvements but cannot directly verify hidden hygiene practices. By reducing cost barriers and providing additional information on potential quality to the consumer, we aim to shift vendors toward a higher-quality equilibrium while reducing moral hazard in a setting with minimal regulatory oversight. Drawing on our consumer findings, our intervention is based on the assumption that maintaining cleanliness is a profitable strategy, and, as vendors transition towards this new equilibrium, increased profitability should enable them to sustain higher-quality operations.

To guide our experimental design, we collaborated with FAO's Food Safety Unit to develop a clear, measurable, and context-specific definition of quality and upgrading appropriate for our setting. Throughout the study, we use these terms to refer to improvements in food safety—a credence attribute that markets may reward but that consumers cannot easily verify at the point of purchase. Because "quality" is otherwise abstract, we began by identifying the most salient contamination risks and the hygiene practices needed to mitigate them, drawing on the food-safety framework outlined in the *Codex Alimentarius* (see, e.g., Codex Alimentarius Commission, 2022).[3] This process informed a set of concrete, observable indicators of "best behavior," distinguishing between the appropriate use of infrastructure and compliance with food-safety protocols. Our measurement focused on three critical and observable domains: water, sanitation, and waste disposal. To promote consistency and reduce measurement bias, we trained enumerators using reference images that defined what counts as "clean and professional" in the local context. This approach ensured that our data collection remained practical for informal vending environments while staying aligned with internationally recognized food-safety standards.[4]

Building on this definition, we designed two treatment packages to address key contamination risks. In the first treatment arm ("T1"), we provide each vendor with two forms of support. The first is a set of "large" visible equipment, including a water storage drum, a stainless steel drinking water container, a hand-washing basin with a fitted water tank, and an 80-liter waste bin. This equipment, delivered at the start of the experimental period, requires minimal effort to use and is highly visible to consumers as it occupies substantial space within the kiosk premises. The second is a set of "small" items essential for safe food practices, such as soap and chlorine tablets for treating water used in kiosk operations. These items are initially delivered weekly at no cost; however, after the treatment period is over (the first three months of the intervention), the free provision ceases, and vendors are instructed on how and where to purchase them to maintain cleanliness. Unlike the larger equipment, these items are less noticeable and require more effort for consistent use. Vendors in the second treatment arm ("T2") additionally receive short, 15-minute weekly training sessions at their kiosks, aimed at improving food safety practices. To track changes in outcomes, we collect two rounds of endline data, along with weekly monitoring data throughout the intervention and for five months afterward. We assess changes in vendors' operations, particularly their use of the provided items and compliance with recommended hygiene practices. This allows us to determine whether short-term improvements persist after subsidies end and whether consumer perceptions align with actual changes in kiosk safety. Overall, our approach represents

[3] The *Codex* does not itself define "food quality." FAO adopts a consumer-oriented notion: "the attributes of a food that influence its value and that make it acceptable or desirable for the consumer" (FAO, 2024), which extends beyond safety alone.

[4] As we explain in detail later, we did not collect food samples for microbiological testing, as this would have required strict on-site protocols that were impractical; and we did not track consumers' health outcomes, which were beyond the scope of this study.

a practical attempt to address low-quality production in an environment characterized by high uncertainty and weak enforcement.

Our main experiment yields several key findings. First, during the treatment period (the first three months), treated vendors consistently use approximately 50% of the large equipment provided and are significantly more likely to use the small items compared to the control group.[5] However, the use of both large and small items declines over time. In particular, while compliance with chlorine tablet use is nearly universal during the initial three months, usage drops to pre-treatment levels—close to zero—once the free provision ends. Second, we observe modest effects on broader kiosk operations: treated vendors are slightly more likely to maintain a cleaner, more hygienic environment and engage in better food handling practices. Third, we find that training has minimal additional impact on equipment use or overall kiosk hygiene. This may be due to vendors perceiving the training as unhelpful: while initial engagement is high (around 90%), participation drops to nearly 5% midway through the program. Fourth, we see significant improvements in business outcomes for the average treated vendor. Equipment provision led to a 7% increase in the number of customers, and 5.7% increase in monthly profits, which corresponds to an average increase of 3.1% relative to the value of the equipment provided. A simple back-to-the-envelope calculation tells us that it would take approximately 809 days to recoup the investment. These findings suggest that improved hygiene practices can yield financial benefits, consistent with our earlier results on consumers. However, we find no evidence that the intervention influences prices, as vendors appear hesitant to change prices relative to their competitors. Fifth, we identify consumer-side externalities in the form of negative local spillovers. Vendors located near a greater number of treated peers exhibit worse business outcomes. Combined with flat prices, this pattern points to reallocation of demand rather than expansion of total market size.

The combined results indicate that while visible, costly equipment enabling quality improvements is valued by consumers and can enhance vendor profits, it is insufficient to motivate sustained changes in vendor behavior. This finding is consistent with a simple model of moral hazard, wherein the perceived long-term returns to cleanliness are too low to sustain behavior change.

To understand this further, we designed a second endline survey to capture potential factors that might explain this outcome. Two explanations stand out. First, strong social pressures within the vendor community sharply limit entrepreneurial choices. Around 70% of vendors report that price and menu coordination is common practice, and 85% agree that those who deviate from the equilibrium risk disapproval from their peers. As a result, even when vendors recognize the potential benefits of differentiating, their scope for doing so is narrow, limiting their ability to recoup a potential investment. Our result is consistent with evidence that collusive social norms in decentralized markets suppress profitable competitive deviations and limit firms' ability to scale (e.g., Breza et al., 2019; Banerjee et al., 2024). Consequently, profits in our market can only increase by expanding sales volumes rather than by changing what is sold or how it is priced. Yet this extensive margin is hard to grow: enlarging a kiosk is prohibitively costly, and vendors already work close to 80 hours per week, leaving little room to extend operating time. Second, vendors operate in a precarious local environment that further dampens incentives for upgrading. Although 75% of vendors express interest

[5] This aligns with baseline reports indicating that 74% of vendors are interested in quality upgrades, stating they would allocate an unexpected financial windfall primarily to equipment maintenance and repairs.

in equipment improvements, 93% cite theft as a major concern. None of the vendors in our sample hold a license to operate, and bribe payments, harassment, and eviction threats are pervasive. The location of kiosks in high-traffic areas with limited access to basic municipal infrastructure further increases the effort required to maintain sanitary conditions. Together, these social and environmental constraints reduce the expected returns to sustained quality upgrading.

Our primary contribution is to the literature on quality upgrading in markets with information asymmetries in emerging economies. Building on evidence that small supply-side subsidies and demand-side incentives are complementary (e.g., Bold et al., 2022; Hoffmann et al., 2023; Park et al., 2023; Deutschmann et al., 2023),[6] we extend this lens to informal street food markets, a complex environment where safety is a credence attribute, transactions are numerous, frequent, and low-value, regulation is weak, monitoring is severely constrained, consumers rely on visible cues, and credible pre-purchase incentives are hard to implement. With respect to prior work, we introduce three key novelties that build on one another. First, we combine demand-side evidence from a consumer DCE with supply-side evidence from a vendor RCT to link the observable side of quality to consumer perceptions and vendor behavior, yielding an integrated view of demand and supply in this market.[7] Second, to make safety improvements visible to consumers while lowering production costs for vendors, we pair *substantial* cost reductions with *visible* equipment upgrades to test whether these interventions can shift equilibrium behavior. Third, we embed an experimental design that measures spatial spillovers, so that returns to upgrading can be interpreted within the local competitive landscape rather than in isolation. While we observe significant short-term effects, vendors have limited incentives to sustain quality improvements. We document contextual factors prevalent in informal markets that weaken the incentive to maintain food safety once external oversight is removed, suggesting that quality upgrading is unlikely to emerge organically in such complex market environments. A key, novel insight from our setting is the prevalence of informal price-fixing norms among vendors, which constrains individual entrepreneurship.

Additionally, we contribute to the emerging economics literature on food safety by documenting poor hygiene practices and inadequate equipment among street food vendors in low- and middle-income countries. Earlier work in this area has been led primarily by food scientists, who have documented the scope and nature of these challenges (e.g., Vollaard et al., 2004; Choudhury et al., 2011; Cortese et al., 2016; Samapundo et al., 2016). More recently, economists have begun to examine food safety in these contexts, providing new insights into behavioral drivers and the impacts of interventions. Our paper is closely related to Daniele et al. (2021), who show that providing vendors with information on safe food practices improves knowledge and awareness, but has limited impact on actual behavior. It also relates to Hoffmann et al. (2023), Ambler et al. (2024), and Cook et al. (2025). The latter two studies, in particular, combine training with the provision of visible equipment to enhance food safety. While they report significant improvements in vendor practices and positive consumer

[6]These studies show that market access and demand incentives are crucial for scaling quality improvements in developing countries. For instance, Bold et al. (2022) and Park et al. (2023) demonstrate that when farmers were linked to buyers offering price premiums for higher-quality produce, their quality-enhancing practices improved substantially. Similarly, Hoffmann et al. (2023) finds that small-scale supply-side interventions, such as subsidies for food safety technology, generate significant short-term improvements when paired with demand-side incentives.

[7]On the demand side, we examine how sanitation-related equipment functions as a quality signal for consumers. Consistent with Bai (2024), we show that investing in visible quality improvements yields positive economic returns by increasing consumer demand. However, unlike this author, we find that these improvements are not sustained beyond the intervention period, as food safety practices decline once external support is withdrawn. Moreover, in line with Michelson et al. (2021), we find evidence of a quality inference problem: because street food safety is difficult to assess directly, consumers rely on observable cues—particularly the condition of a vendor's equipment—to infer overall food quality, even though these signals do not necessarily reflect actual hygiene practices.

responses to increased cleanliness, they find minimal changes in the microbiological quality of the food.[8]

The rest of this paper is organized as follows. Section 2 documents our baseline assessment of the setting. Section 3 contains the consumer-side experiment. Section 4 describes the experimental design for the vendor-side intervention and the data. Section 5 presents the estimation strategy, the results, and delves into potential mechanisms. Section 6 outlines a simple model of vendor decision-making under moral hazard with imperfect monitoring to help explain our results. Section 7 concludes.

2 Context

We focus on three areas in Kolkata, India; namely, Dalhousie, Hazra, and Sector V. These areas have a large number of street vendors who prepare and cook food at the kiosks and typically handle items with a higher risk of contamination and require more extensive equipment for food preparation.[9] In the Appendix, Figure A1 shows the locations of these areas in the city and Figure A2 shows a picture of the typical vendor in our sample. Using data collected at baseline, we document several key facts that confirm that the quality of street food produced from a food safety perspective is low (e.g., Daniele et al., 2021).

We begin by describing the main operational issues reported by vendors in our sample that could potentially serve as barriers to quality upgrading. The most pressing problem identified is the lack of essential sanitary infrastructure for safe food preparation, with more than 40% of vendors reporting problems with access to toilets and potable water (Table 1). Equally concerning are the significant hurdles vendors face related to the local institutional environment. 43% of vendors report regulatory uncertainty, bribes, and the lack of proper licenses and permits as major concerns. None of the vendors in our sample held any official license during the time of our study, leaving them vulnerable to bribery and extortion in exchange for permission to operate. Theft is also likely an issue: kiosks are typically only fastened with padlocks at night, and there is limited police surveillance of kiosk areas. Other key challenges include the lack of electricity (27%), competition from other vendors or formal businesses (24%) and difficulties related to cost or access to finance (19%).

The bottom section of Table 1 provides an overview of baseline kiosk facilities and inputs related to food safety.[10] While almost all vendors report having some form of drinking water facility, hand-washing station, and garbage bin, qualitative data reveal that these facilities are often rudimentary and improvised.[11] In addition, basic hygiene items are inconsistently available: hand-washing soap is often missing, aprons are rarely used, and few vendors comply with hand hygiene practices or hair covers. Chlorination of water is virtu-

[8] Tangentially, we also contribute to the literature on the role of subsidies and information in influencing health-related behaviors. Regarding subsidies, we find that providing free or heavily subsidized health-related products significantly increases short-term adoption, consistent with the findings of Kremer and Miguel (2007), Cohen and Dupas (2010), and Dupas et al. (2016). However, we observe a decline in product usage once subsidies are withdrawn, indicating a low willingness to pay for these products post-intervention. This mirrors patterns identified in Ashraf et al. (2010), Kremer et al. (2011), Blum et al. (2014), and Ritter et al. (2017). Our findings also resonate with broader research showing that while subsidies can effectively boost initial take-up, they often fail to result in long-term behavioral change without sustained support (Hanna et al., 2016; Fischer et al., 2019; Shukla et al., 2022). Turning to information-based interventions, while some studies have found positive effects (e.g., Rhee et al., 2005; Ashraf et al., 2010; Devoto et al., 2012; Luoto et al., 2014), our results show no significant impact of training on equipment usage or hygiene practices. These findings align with research suggesting that simply providing information about health products or safe behaviors does not necessarily translate into meaningful behavioral change (e.g., Kremer and Miguel, 2007; Dupas, 2009; Nyhan et al., 2014; Duflo et al., 2019; Ho et al., 2023).

[9] Given our interest in the role of equipment and training, we do not focus on street food vendors who sell drinks or cold snacks, and who prepare food at home rather than at the kiosk. More information on our sample of vendors is provided in Section 4.

[10] Table A1 in the Appendix breaks down this information by study area.

[11] The qualitative assessment of asset quality was conducted during an initial round of field visits, with data recorded on notepads. These statistics are not included in the table because nearly all observed vendors had very poor-quality inputs.

Table 1: Initial Context Assessment: Problems and Facilities

	Obs.	Mean	S.D.	Min	Max
Main problems encountered during normal business operations:					
Access to toilets	274	0.42	0.49	0	1
Access to potable water	274	0.40	0.49	0	1
Access to electricity	274	0.27	0.44	0	1
Bribes, licensing, permits	274	0.43	0.50	0	1
Cost or access to finance	274	0.19	0.40	0	1
Competition from others	274	0.24	0.43	0	1
Facilities and inputs:					
Kiosk has handwashing facility	508	0.95	0.22	0	1
Kiosk has garbage bin	508	0.88	0.33	0	1
Kiosk has drinking water facility	508	1.00	0.06	0	1
Handwashing facility has soap	483	0.43	0.50	0	1
Vendor uses an apron	508	0.08	0.28	0	1
Vendor wears gloves	508	0.00	0.04	0	1
Vendor wears hair cover	508	0.01	0.11	0	1
Treats water in primary storage	508	0.05	0.21	0	1

Notes: Data are pooled from two pre-treatment surveys (surveys 1 and 2) conducted during our initial assessment of the context in May 2022. See Section 4.4 for details on the data collection. Each variable is binary taking value 1 if the item or behavior is observed, 0 otherwise.

ally absent: only 5% of vendors report that they chlorinate the water used in their daily kiosk operations.[12] Figure A3 provides photographic examples of the baseline conditions under which vendors in our study operate. These images illustrate the challenging environment: food preparation often occurs in close proximity to garbage and stagnant water, with little to no equipment available for washing dishes or hands.

A major input to street food kiosk operations is water, which is used for cooking, drinking, and cleaning. In our setting, no kiosk has direct access to a water source. Instead, most vendors collect water in large containers from a local public tap, either for free or with a charge, in the morning and transport it to the kiosk.[13] To ascertain the quality of the water used by vendors in their kiosk operations, we randomly sample 25 water sources that vendors in our study area.[14] We collect four samples from each source over the course of four months and test the samples for: (i) total coliforms, which provide an overall indication of the bacterial condition of the water and indicate the presence of pathogens;[15] (ii) Escherichia coli (E. coli), a subset of total coliforms and a key indicator of fecal contamination;[16] and (iii) total bacterial counts, a complementary indicator to coliforms. Table 2 lists the results. Total coliforms is detected in 69% of the water samples we collect and E. coli is detected in 45% of samples. Total bacterial counts are also extremely high, with an average of almost 4,000 CFUs/mL. 54% of water samples have more than the recommended maximum count

[12] This aligns with findings from other developing country contexts. For example, Nizame et al. (2019) report that only 11% of street food vendors in Dhaka, Bangladesh, had soap and water for hand-washing, while Samapundo et al. (2015) note that 60% of kiosks in Port-au-Prince, Haiti, had visible flies and animals nearby, and 65% lacked access to potable water.

[13] Kiosks have a substantial water requirement, with vendors reporting an average usage of around 174 liters per day. In areas with limited access to local taps, like Sector V, vendors often choose to purchase water from a local supplier who collects it from public taps and delivers it directly to the kiosk.

[14] We select 10 in Dalhousie and Sector V, and 5 sources in Hazra. We also test the water that the vendors have in their kiosks directly for the presence of chlorine; more information is provided in Section 4.5.

[15] Note that the presence of coliforms does not necessarily indicate unsafe water; however, the recommended level of total coliforms in potable water is zero (FAO, 2023).

[16] E. coli is naturally found in the intestines or humans and animals, and is a widely used indicator for detecting fecal contamination. While not all strands of E. coli are harmful, many strands cause diarrhea and vomiting, and can lead to respiratory illness or pneumonia.

of 500 CFUs/mL.[17]

Table 2: Bacterial Quality of Local Water Sources

	(1) Dalhousie	(2) Hazra	(3) Sector V	(4) Total
Laboratory analysis:				
Total coliform detected [0,1]	0.68	0.50	0.80	0.69
E.Coli detected [0,1]	0.55	0.40	0.38	0.45
Total bacteria counts (CFUs/mL)	4413	3573	3368	3827
CFUs/ml>500 [0,1]	0.53	0.50	0.57	0.54
Obs.	40	20	40	100

Notes: The data were collected by a local independent inspection and testing company in Kolkata (Mitra S. K. Private Limited) between October 2022 and January 2023. Each column reports the mean value. Water sources are public, typically taps located in the city. Total coliforms indicates the overall bacterial quality of the water, with the recommended level in potable water being zero. E.coli is a subset of total coliforms, and CFU stands for Colony Forming Units. The recommended total bacterial counts for potable water is less than 500 CFU/mL.

In addition to problematic infrastructure and inputs, knowledge and practices regarding safe food preparation and handling are also an issue.[18] Table 3 provides descriptive statistics for outcomes related to various food safety practices collected at baseline. While some practices are widely observed (using detergent for washing dishes, using a towel or cloth for hands, and using tongs or spoons during food preparation), others see far lower compliance. Only 62% of vendors were found to be using clean dishwater, and just 40% had a garbage bin that was clean and empty. Less than 60% of vendors protect cooked food from potential contaminantion by covering it or placing it behind a screen, and only 38% wash their hands with soap before handling food. The data also shows that just 21% of vendors ensure that raw food is fully covered, and only 21% use disposable plates, which can mitigate cross-contamination risks.

We also collect census information on all items sold by vendors, including menu composition and prices. Vendors sell an average of six items. Meals typically consist of rice, noodles, or roti served with vegetables, eggs, or chicken, as well as platters (thali) that include vegetable, chicken, or fish options. The average meal costs 42₹, with prices generally higher in Sector V and lower in Dalhousie. To explore how prices relate to vendors' resources and practices at baseline, we run item-level regressions using two dependent variables: an asset index (based on the facilities and inputs listed in Table 1) and a food-safety practices index (based on the variables in Table 3), both averaged across our two pre-treatment surveys. We find no significant relationship between item prices and either of these measures (see Table A2 in the Appendix).

To conclude our assessment, we gathered information about vendors' desire to upgrade the quality of their kiosks. When asked how they would invest a hypothetical 10,000₹ windfall in their businesses, 73% of ven-

[17] The fact that the water quality is poor water quality is likely to be known by residents of Kolkata. Media reports indicate that the use of water filters or purifiers to treat water is very common among urban Indian households; see e.g., The Times of India (2019); The Telegraph India (2024); The New Indian Express (2023). The reasons for poor water quality are numerous and include aging and leaky pipelines that are susceptible to cross contamination (Singh et al., 2015) and untreated sewerage and industrial contaminates that are discharged directly into water sources (Central Ground Water Board, 2022-2023). Effective chlorination, a fundamental part of water purification, is often challenging for municipalities. Part of this is due to the variable quality of water, such that it is difficult to maintain optimal chlorine dosages for disinfection. Leaky pipelines may mean that water at one point of the system is potable but then becomes contaminated at another point. Intermittent water supply, a key problem in Kolkata, is also known to cause contamination (Bivins et al., 2021; Satpathy and Jha, 2022).

[18] Poor food safety practices among street food vendors in many low- and middle-income countries have been well-documented (e.g., Vollaard et al., 2004; Choudhury et al., 2011; Cortese et al., 2016; Samapundo et al., 2016; Daniele et al., 2021).

Table 3: Initial Context Assessment: Food Safety Practices

	Obs.	Mean	S.D.	Min	Max
Uses soap for dishes	344	0.93	0.26	0	1
Dish water is clean	412	0.62	0.49	0	1
Garbage bin is clean and empty	447	0.40	0.49	0	1
Uses disposable plates	508	0.21	0.41	0	1
Counter is clean	508	0.69	0.46	0	1
Uses towel/cloth for hands	508	0.97	0.16	0	1
Cooked food is covered or behind screen	508	0.56	0.50	0	1
Raw food is fully covered	354	0.21	0.41	0	1
Uses tongs or spoons	508	0.76	0.43	0	1
Washes hands with soap	496	0.38	0.49	0	1

Notes: Data are pooled from two pre-treatment surveys (surveys 1 and 2) conducted during our initial assessment of the context in May 2022. See Section 4.4 for details on the data collection. Each variable is binary taking value 1 if the food safety practice is observed, 0 otherwise. Note that "Uses soap for dishes" refers to the presence of soap in the dish water, "Counter" refers to the main counter where food is prepared in the kiosk, and "Washes hands with soap" refers to washing hands with soap before handling food. Missing observations stem predominately from the data collectors being unable to observe the practice.

dors indicate that they would prioritize spending on items related to the maintenance, repair, and renovation of their kiosks or the buying or upgrading customer service equipment (see the top panel of Table A3 in the Appendix). This response was far more common than those related to expanding the size of the kiosks through either new food items or hiring more workers. This aligns with our earlier observations of the rudimentary equipment currently in use at most kiosks. However, despite their recognition of and desire for quality improvements, credit constraints may pose a significant barrier. Indeed, although 93% of vendors report having a bank account and 57% of them use this account for their business operations, borrowing remains uncommon, with only 16% of vendors having ever applied for a business loan (bottom panel of Table A3 in the Appendix).[19] Furthermore, 53% of vendors report having no savings in their bank accounts or at home, and the average monthly savings in the sample is 2,739₹ (around 33 USD).

3 Consumer Preferences for Safe Food

In this section, we explore what consumers value in street food and their willingness to pay for safer options. While most consumers presumably prefer safer and cleaner food, all else being equal, assessing food safety can be challenging. Cues such as kiosk cleanliness and the vendor's hygiene likely provide signals about the safety of the food and consumers who prioritize safety may therefore value food from a clean kiosk where vendors appear to follow good hygiene practices. Even if these measures do not fully eliminate contaminants, they help reduce the risk of exposure. Our analysis serves two purposes. First, we examine the demographics and preferences of street food consumers. Second, we implement a Discrete Choice Experiment (DCE) (e.g., WHO, 2012; Mas and Pallais, 2017; Wiswall and Zafar, 2018; Daniele et al., 2021; Maestas et al., 2023) to quantify how much consumers are willing to pay for perceived improvements in food safety. This allows us to estimate the relative importance of various attributes when consumers choose among street food options,

[19]Furthermore, fewer than 10% of vendors have applied for a loan through a microfinance institution or informal lender (not reported in the table). We do find that upon applying for a loan, most vendors report success in their application (75% for a formal bank loan, 92% from a microfinance institution, and 94% from an informal lender). At the time of our baseline survey, 10% of vendors currently had a bank loan, 8% had a microfinance loan, and 3% had a loan from an informal lender.

including the trade-off between price and cleanliness.

3.1 Who Consumes Street Food?

We collected survey data from 2,684 repeat consumers purchasing street food across the three areas of interest: Dalhousie, Hazra, and Sector V. Our focus was on consumers who were regular customers of vendors identified for the supply-side intervention, with the goal of surveying approximately 10 consumers per vendor. Sampling was conducted during peak hours, between 11 a.m. and 4 p.m. Table 4 presents descriptive statistics, both by area and in aggregate. Customers tend to be predominantly male and in their mid to late 30s. On average, around 44% of consumers hold a college degree, and approximately 32% report being employed in high-skilled occupations. Sector V, the city's IT hub, has a higher concentration of younger, college-educated, and high-skilled consumers. Nearly three-quarters of respondents report consuming street food several times per week or more.

Table 4: Consumer Descriptive Statistics

	Dalhousie (1)	Hazra (2)	Sector V (3)	Total (4)
Demographics:				
Male	0.89	0.77	0.82	0.85
Age	37.52	36.88	30.62	35.60
Has college degree	0.39	0.37	0.61	0.44
Is high-skill employee	0.30	0.20	0.45	0.32
Is low-skill employee	0.34	0.23	0.21	0.29
Consumption habits:				
Eat street food frequently	0.75	0.69	0.69	0.72
Find it easy to detect unsafe food	0.41	0.39	0.42	0.41
Has been sick from street food before	0.30	0.33	0.35	0.32
Would change vendor after getting sick	0.53	0.59	0.55	0.54
Important factors when choosing street food:				
Hygiene	0.91	0.86	0.92	0.90
Taste	0.64	0.63	0.77	0.67
Price	0.40	0.53	0.42	0.43
Relationship	0.25	0.22	0.25	0.24
Location	0.11	0.10	0.08	0.10
Health	0.06	0.03	0.05	0.05
Most important determinant:				
Hygiene	0.73	0.72	0.76	0.74
Food safety knowledge:				
What is contaminated food? [0,7]	1.87	1.91	1.59	1.80
Obs.	1464	520	700	2684

Notes: Data from consumer survey. The table provides average values for a range of variables by area and for the full sample. With the exception of the variables "age" and the "food safety knowledge", all variables are indicator variables [0,1]. For the question on "important factors" when choosing street food, respondents could select multiple answers. For the "most important determinant," respondents were asked to choose a single answer from the same set. Food safety knowledge is an index equal to the average number of correct responses to a question asking respondents to list common sources of food contamination. Answers were aggregated into 7 broad categories.

One important issue from the consumer perspective is that contamination is largely invisible at the point of purchase. When asked, *"In your experience, is it hard for you to tell whether the street food you eat is safe or not?"*, only 41% of respondents said that it is easy to identify unsafe food.[20] Regarding past experience, 32% report having fallen ill after consuming contaminated street food at some point. However, just over half of consumers state that they would switch vendors if they became sick from food purchased at a particular kiosk. This suggests a potentially moderate preference for food safety. At the same time, 90% of respondents say that vendor and kiosk hygiene is an important factor in their decision to choose where to eat—more than taste (67%), price (43%), or location (10%). When asked to identify the most important factor in their decision, 74% of consumers selected hygiene. On the other hand, consumer knowledge about food safety appears relatively low: respondents correctly identified, on average, only 1.8 out of 7 possible food contaminants listed in our questionnaire.[21]

3.2 Discrete Choice Experiment

We utilize two hypothetical street food items in our DCE: vegetable thali and chicken thali. These items were chosen because they are relatively homogeneous and easily recognizable to consumers.[22] Each consumer was randomly assigned one of the two food items and presented with 18 binary choice scenarios, where they selected between option A and option B. The order of scenarios was randomized across consumers. Each choice varied across four attributes: three with two levels, and one (price) with five levels.[23] The attributes and their levels were as follows:[24] (i) kiosk's hygienic conditions: "appears very clean and hygienic" vs. "appears not very clean and hygienic;" (ii) vendor's personal hygiene: "appears very clean and hygienic" vs. "appears not very clean and hygienic"; (iii) location: "vendor is in front of you" vs. "vendor is a 5-minute walk from you (about 400 meters)"; and (iv) prices: 30, 40, 45, 50, and 60₹ for the vegetable thali, and 50, 70, 80, 90, and 110₹ for the chicken thali. To help respondents differentiate between "very clean and hygienic" vs. "not very clean and hygienic," we provided example images. Price levels were selected to reflect typical market ranges for each item. An example choice scenario is shown in Figure A4 in the Appendix.

We focus on these attributes for two main reasons. First, the two cleanliness attributes help identify which hygiene-related dimension—kiosk conditions or vendor appearance—matters more to consumers. This distinction also directly maps to our supply-side interventions: improving kiosk hygiene through equipment, and improving vendor hygiene through training. Second, the location attribute captures how much consumers value proximity when choosing street food. More importantly, it offers indirect evidence on the potential for consumer-driven spillovers in our experiment. While vendors in different clusters are physically separated, consumers may be more mobile. If consumers are willing to walk farther to access cleaner or safer options,

[20] This item captures consumers' perceived ability to judge safety. In practice, most foodborne hazards cannot be reliably detected by sight, smell, or taste; proper assessment requires trained inspection of risk practices and, where necessary, laboratory testing. We include this question to document consumer beliefs and perceived control, rather than to measure actual ability to detect contamination.

[21] This figure comes from an open-ended question: *"What do you think can make food unsafe to eat?"* Enumerators recorded verbatim answers, which we coded into seven categories. The statistic is the average number of distinct categories named. Regarding microbiological risks, only 8% of consumers explicitly mentioned microbiological contamination.

[22] Moreover, the vegetarian option carries a relatively lower food safety risk compared to the chicken option. Hence, we can use the two items as a robustness check: if consumers value food safety, they should be willing to pay more for safety improvements in the item with higher contamination risk.

[23] We followed best practices in DCE design, using a statistically efficient, D-optimal design to select choice sets. D-efficiency minimizes the determinant of the covariance matrix, thereby reducing the standard errors of parameter estimates.

[24] A vendor who "appears very clean and hygienic" is described as having clean hands and clothes, wearing a clean apron and head covering, and operating in clean surroundings. A kiosk that "appears very clean and hygienic" includes a professional handwashing station, clean drinking water, a clean cooking area, and a dustbin. Enumerators were trained with reference photos illustrating these conditions, and respondents were shown the same images during the DCE to anchor the cleanliness descriptions. Additional materials, including the images, full questionnaire, and enumerator manual, are available upon request.

treatment effects in the vendors experiment could reflect a reallocation of customers from untreated to treated vendors rather than a pure improvement in the safety and hygiene of food offered by the same vendors. If, instead, proximity strongly affects consumer choice, demand is likely localized—reducing concerns about such spillovers. We come back to this point in Section 5.5.

3.3 Consumer Willingness-to-Pay for Safer Street Food

To estimate Willingness-to-Pay (WTP) using the DCE data, we specify a utility function for each alternative in the choice set and estimate its parameters using a mixed logit model, which maximizes the likelihood of the observed choices given the utility specification. In this model, we allow the parameters associated with the attributes to vary across consumers. Once the model is estimated, we compute the expected WTP for each attribute as the ratio of its coefficient to the price coefficient. This ratio reflects the average amount a consumer is willing to pay for a one-unit increase in a given attribute. Formally, if β^a is the coefficient for attribute a and β^{price} is the coefficient for price, then the WTP for attribute a is given by $-\frac{\beta^a}{\beta^{\text{price}}}$. Further details are provided in Appendix B.

Table 5: Consumer Willingness to Pay by Attribute

	(1) Veg item	(2) Non-veg item	(3) Full sample
Clean kiosk (₹)	79.7***	96.1***	90.9***
Clean vendor (₹)	29.1***	30.6***	30.7***
Location > 5 min walk (₹)	-2.9***	-2.0***	-2.5***
Obs.	48,312	48,312	96,624
Avg. price in the market (₹)	32.9	73.4	51.8
Avg. price in non-AC restaurant (₹)	130-150	180-200	150-170
Avg. price in AC restaurant (₹)	160-180	230-250	200-220
Avg. price in luxury AC restaurant (₹)	230-250	320-350	280-300

Notes: Data from consumer survey. WTP estimates are derived from a mixed logit model that allows for random coefficients across consumers. Full parameter estimates are reported in Table A4 in the Appendix. The top panel reports implied WTP for safer street food options in Indian Rupees (₹). The bottom panel reports average prices for the same meals in the street food market in Kolkata and in restaurants. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5 lists our WTP estimates in rupees (₹) and benchmarks them against prevailing street and restaurant prices.[25] The results reveal a clear consumer preference for cleanliness, especially at the kiosk level. First, the average WTP for a clean kiosk is 91₹, with slightly higher values for the non-vegetarian meal (96₹) than for the vegetarian meal (80₹). These represent substantial premiums—nearly three times the market price of a vegetable thali (33₹) and over 30% above that of a chicken thali (73₹). Second, consumers also value the vendor's personal hygiene, though to a lesser extent. The average WTP is approximately 31₹, which is nearly equal to the price of a vegetable thali and about 40% of the price of a chicken thali. This suggests that both vendor appearance and kiosk infrastructure serve as meaningful signals of food safety. Third, location has a much smaller effect, indicating that distance is not a major factor in consumer decision-making. The small disutility from walking implies that consumers may be willing to travel short distances to access cleaner

[25]In Appendix, Table A4 presents the estimated coefficients for each attribute using our mixed logit model.

food, raising the possibility of demand-side spillovers in the vendor experiment. Overall, even the highest WTP estimates remain below the average price of comparable meals in non-AC restaurants (130–150₹ for vegetarian items, 180–200₹ for non-vegetarian), reinforcing the idea that cleaner street food is seen as a safer yet still affordable alternative.

These results are robust to a range of checks. First, we included a simple rationality test within the DCE to ensure that consumers were actively engaging with the choice tasks. Specifically, two of the eighteen scenarios were constructed to feature a clearly dominant option (offering lower price, better kiosk and vendor hygiene, and closer proximity), and were randomly placed within the sequence. Nearly all respondents selected the dominant option in both cases, suggesting that choices were deliberate rather than random. Second, we examine how WTP correlates with observable characteristics using individual-level estimates from the mixed logit model. Table A5 in the Appendix presents OLS regressions of WTP measures on selected covariates. Respondents who report hygiene as the most important factor display significantly higher WTP for both kiosk and vendor cleanliness. We also find positive and significant associations between WTP and socioeconomic status, particularly education and high-skill employment, especially for kiosk hygiene, the attribute consumers rated as most important. Third, we assess the sensitivity of our results by re-estimating the model using a conditional logit specification. The results, shown in Table A6 in the Appendix, are qualitatively similar to those from the mixed logit model. Finally, we use the conditional logit model to explore heterogeneity in WTP by area and demographic group. Table A7 shows substantial geographic variation: WTP for kiosk hygiene is highest in Sector V and lowest in Hazra, consistent with differences in consumer profiles across locations. Table A8 highlights clear socioeconomic gradients: high-skill consumers are willing to pay considerably more for both kiosk and vendor cleanliness. These patterns are consistent with the idea that higher-income individuals place greater value on hygiene and are more selective in their food choices.

To put our WTP estimates in perspective, consider a simple back-of-the-envelope calculation. Our intervention (described later) provided each vendor with equipment worth about 29,117₹. If a vendor could fully pass through the average WTP premium for a cleaner kiosk ($\approx$ 91₹ per meal) without losing any demand, about 320 meals would be enough to recover this investment; whereas, using only the premium for vendor hygiene ($\approx$ 31₹) would require around 940 meals. Given that vendors typically sell 70 meals per day, this implies up to 16 days of sales to recoup the value. While this is an intentionally optimistic upper bound (it assumes perfect price pass-through, no drop in demand, and treats each vendor in isolation, ignoring competitive reactions and potential customer reallocation), it still fits the point emphasized by Hoffmann et al. (2019): stated-preference WTP should not be interpreted as literal price premia. Its main value lies in revealing direction and relative importance rather than predicting actual market prices. Thus, our consumer experiment should be viewed primarily as evidence that hygiene matters to consumers, which is sufficient to motivate the supply-side intervention that follows.

4 Field Experiment with Vendors

Our previous experiment showed that consumers are willing to pay significantly more for meals from cleaner-looking kiosks and vendors perceived as hygienic. However, that evidence was based on stated preferences, not actual behavior. We now translate these insights into practice through a randomized intervention with street vendors, evaluating whether a substantial capital subsidy can shift vendors toward higher food safety standards. To inform the intervention, we collaborated with a team at FAO's Food Safety Unit to first develop a context-specific definition of quality in terms of food safety. We focus on features that are both critical for reducing contamination risks and visible to consumers. The program provided treatment group vendors with sanitation-related equipment, hygiene supplies, and, for a subset of vendors, food safety training. The core idea is simple: if vendors can upgrade at low cost and consumers can recognize those improvements, then maintaining cleanliness should become a profitable and self-sustaining strategy.

4.1 Defining and Measuring Quality

In collaboration with FAO, we developed a context-specific, measurable definition grounded in science-based principles from the *"Codex Alimentarius"* (Codex Alimentarius Commission, 2022) and tailored to the informal street food sector.[26] Our definition distinguishes between two core dimensions: (i) the presence and condition of key sanitation-related *inputs*, and (ii) compliance with observable food safety *practices*—both of which collectively reduce the risk of contamination during food preparation and service. Each of these dimensions is directly targeted by our intervention and monitored throughout implementation.

In terms of inputs, we focus on some of the main physical infrastructure required to support safe food production, including equipment and supplies that enable access to clean water, hand-washing, and waste disposal, as described in detail in Section 4.3. For practices, we identify hygienic behaviors across six domains—personal hygiene; equipment and utensils; control of operations; waste management; cleaning and sanitizing; and raw materials—as shown in Figure A9 in the Appendix. Within each domain, we selected a set of behaviors that could be reliably observed during monitoring visits. These are marked with green check-marks in the Figure and were operationalized into a set of structured monitoring questions with predefined response categories. The full list of items is available in Tables A9 and A10 in the Appendix, and forms the basis of the outcome indices described in Section 5.1.

Two important limitations of our measurement strategy should be acknowledged. We did not collect food samples for laboratory testing to assess the microbiological safety of meals. In fact, our donor strongly advised against it for two main reasons. First, reliable microbiological analysis requires rigorous on-site sampling and immediate testing; without such protocols, results could be misleading and difficult to link to our intervention, given the many factors that can cause contamination. Second, visible sampling in the streets would likely have attracted unwanted attention, created disruption, and discouraged vendors from participating. Similarly, we did not collect data on consumers' health outcomes, as this was beyond the scope of our experiment. Linking vendor-level interventions to both microbiological measures and consumer health remains an important avenue for future work aiming to assess food safety more directly.

[26] The term Codex Alimentarius is Latin for "food code." The Codex Alimentarius consists of international food standards, codes of practice (including hygienic codes), guidelines, and other recommendations designed to protect consumer health and ensure fair practices in the food trade. The collection of these standards and related texts, adopted by the Codex Alimentarius Commission, is known collectively as the Codex Alimentarius. The Codex Alimentarius Commission (the "Commission") was established by the Food and Agriculture Organization (FAO) and the World Health Organization (WHO) under the Joint FAO/WHO Food Standards Programme. It coordinates input from 189 member countries to develop and endorse the international food standards that make up the Codex Alimentarius. The Codex Alimentarius is accessible on the official website here.

4.2 Experimental Design

Our sample is restricted to street food vendors who meet the following criteria: (i) they cook and sell food at their kiosks; (ii) they offer meals or lunch/dinner items; and (iii) they provide at least three varieties of food options. After conducting a comprehensive survey of vendors in the three specified areas of Kolkata, we identify 284 vendors who meet these criteria. Our local NGO then acted as an intermediary between our research team and the selected vendors, along with local union leaders, to secure their consent to participate.

In each of the three geographic areas, we categorize the sampled vendors into three types of natural urban clusters based on the number of other vendors within a 30-meter radius. The first set of clusters comprises vendors who have no other sampled vendors within this radius, making them the sole vendor in their cluster. The second set consists of vendors with one or two neighboring vendors within the 30-meter radius. The third set includes vendors with three or more other sampled vendors nearby. In streets where vendors are situated closely together, we measure the distance from the leftmost to the rightmost vendor and consider a cluster to be a collection of vendors located within 30 meters of each other. Approximately 89% of the vendors fall within cluster sizes ranging from one to five vendors.[27]

Three primary considerations guide our clustering strategy. The first is fairness: we aim to avoid giving a vendor an advantage over their immediate neighbor. In a few cases where a control vendor and a treatment vendor were next to each other, we worked closely with local union leaders to ensure that our randomization process did not disrupt social cohesion. The second is to ensure clear separation between treatment and control clusters. The 30-meter distance between vendors aims to minimize contamination of control clusters through vendors directly observing what is happening at another kiosk. However, this does not prevent *consumers* from observing differences across vendors and we address potential consumer-side spillovers in a robustness check later in the paper. The third consideration is the need to account for similar competitive environments and operating conditions, as vendors with nearby competitors may operate their businesses differently from those without close competition.

We conduct a stratified random assignment with two levels. Stratification is done at the area and cluster size level.[28] With three areas and three cluster size categories, this yields nine strata. Clusters are randomly assigned to one of two treatment groups (described below) or the control group. Our randomization yields 40 clusters in the control group (97 vendors), 35 clusters in the first treatment group (94 vendors), and 36 clusters in the second treatment group (93 vendors).

4.3 Details of the Intervention

We provide all vendors in our study with personalized banners for display, should they wish to use them, featuring the kiosk's name and a menu card. All vendors chose to display the banners, which allowed us to easily identify them for monitoring after the intervention. Vendors in both treatment groups ("T1" and "T2") receive essential sanitation-related items, none of which are provided by local authorities, and some of which

[27] The remaining 11% belong to larger clusters. Specifically, clusters with 6 vendors (1 case), 8 vendors (2 cases), and 9 vendors (1 case). Figure A5 in the Appendix shows the distribution of sampled vendors in the three areas, as well as an example of how we define a natural urban cluster on the street.

[28] See, e.g., Imai et al. (2009), Imbens (2011), and Blair et al. (2019) for a technical discussion of randomization at the cluster size level.

are costly to obtain. These items are distributed in June–July 2022,[29] at the beginning of the study period. In one of the treatment groups ("T2"), we also cross-randomize short training sessions on the proper use of these facilities, aiming to address the challenges highlighted in Table 3.[30] No vendors in the treatment groups declined to receive the facilities or participate in the training.

Figure A6 in the Appendix illustrates the structure of the intervention. Both T1 and T2 received "large equipment" and "small supplies." The large equipment consists of durable, highly visible items valued at approximately 350 USD per vendor—roughly equivalent to two months of average profits in our sample (comparable to, e.g., De Mel et al., 2008). These items include: (i) a water storage drum, (ii) a stainless steel drinking water container with a tap, (iii) a hand-washing basin with a fitted water tank, and (iv) an 80-liter waste bin. Most vendors already owned some version of these items, but in many cases the new equipment replaced low-quality or worn-out items, and in others, represented entirely new additions. Vendors had full discretion over how to use this equipment, as no usage conditions were imposed.

The small supplies consist of non-durable, less visible hygiene and sanitation items that were delivered weekly for free over a 12-week period, starting in the second week of July 2023. These include: (i) hand-washing soap, (ii) aprons, (iii) hairnets, and (iv) chlorine tablets for disinfecting water from the vendor's primary source.[31] These supplies required vendors to change daily routines—for example, by treating water each morning and consistently using protective clothing. After the 12-week subsidy period ended, vendors were no longer provided these supplies for free, but received detailed information on how and where to purchase them locally, tailored to their specific locations.

The food safety training component was provided exclusively to T2 vendors over the same 12-week period. Each week, vendors received a brief (15-minute) visit from a trained facilitator. These sessions covered the purpose and proper use of each item in relation to safe food practices, introduced a set of food hygiene rules, and included short interactive exercises to reinforce learning. Trainers followed a structured curriculum with clear objectives and key messages, focusing on both equipment use and general hygiene standards.[32] Each visit was documented using standardized tracking forms, which captured vendor engagement, progress toward specific goals, and whether the equipment was in active use.

4.4 Data Collection

Figure A8 in the Appendix provides a timeline summarizing our data collection. We collected baseline data in April 2022. The treatment period started in July 2022 and lasted for 12 weeks. In September 2022, at

[29]Due to logistical constraints, the delivery of the large equipment was staggered over a three-week period, with the first group of vendors receiving the items at the end of June 2023. As a result, during the first two weeks of the delivery period (period 3 and 4), a number of treated vendors had not yet received the equipment, effectively giving rise to an event-study-style variation in treatment timing. To simplify the econometric analysis and ensure comparability across vendors, our main specification excludes observations from these initial two weeks of monitoring and focuses on post-treatment outcomes starting from the third post-treatment monitoring round (period 5), once all vendors had received the equipment. In a robustness check, we retain the full set of observations and estimate the effects using an event study framework. Results are consistent and quantitatively similar.

[30]We do not include a training-only treatment group because prior work in a similar context, such as Daniele et al. (2021), finds no significant effects of training alone on food safety behavior or business outcomes. Based on this evidence, and considering resource constraints, we judged a training-only group to have limited policy relevance.

[31]Treating water with chlorine tablets is one of the more straightforward and cost-effective ways to make water potable. A large literature has shown that treated water reduces diarrheal disease at the household level; see, for example, Fewtrell et al. (2005), Arnold and Colford (2007), Clasen et al. (2007), and Haushofer et al. (2021).

[32]An excerpt from the booklet with the full set of hygiene rules is provided in Figure A7 in the Appendix. The full booklet is available here.

the end of the treatment period, we conducted our first endline survey (Endline 1). Both surveys gathered socio-economic and business data, including information about vendors' demographic characteristics, household welfare, business practices, business assets, business financing, and vendors' behavior and awareness regarding food safety. In Endline 1, we additionally collected detailed information on how vendors utilized the provided equipment or if they purchased any new equipment. To assess medium-term impacts, we carried out a second endline survey (Endline 2) in February 2023, approximately five months after the end of the treatment period and eight months after the initial delivery of equipment.

To track changes throughout the study period, we collected detailed monitoring surveys. Specifically, we conducted two surveys prior to the delivery of equipment and training (in May 2022), nine weekly monitoring surveys during the treatment period (from July to September 2022), and five biweekly monitoring surveys in the post-treatment period (from November 2022 to February 2023). These surveys aimed to quantify vendor behavior during peak business hours, with a particular focus on safe food practices, such as kiosk cleanliness and vendors' hygienic behavior. This information was gathered through random audits during regular business operations to minimize the risk of self-report bias.[33] At the end of each audit, vendors answered a brief set of questions about their business operations and food safety practices.[34]

To assess both the quality of the water used by each vendor and whether vendors in the treatment groups are using the provided chlorine tablets, we collected individual water samples every two weeks. These samples are then tested for the presence of chlorine. Specifically, we collected 10ml of water from the kiosk's primary water storage container between 12 p.m. and 2 p.m. and analyzed the samples for free chlorine levels using a professional chlorine tester.[35] We collected three pre-treatment water samples and continued sampling throughout and after the treatment period, for a total of up to ten samples per vendor between July 2022 and February 2023.

Attrition during the study was low. Ten vendors dropped out after the baseline due to personal reasons, such as illness or returning to their villages. Consequently, our post-baseline sample consists of 274 vendors, 97 in the control group (40 clusters), 92 in the first treatment (35 clusters) and 85 in the second treatment (33 clusters). During the monitoring surveys, there were instances in which vendors were not present at their kiosks when enumerators visited. However, as demonstrated in Table A11 in the Appendix, we successfully maintained low and non-differential attrition rates during the entire study period, including for Endline 1 (3% attrition rate), and for Endline 2 (7% attrition rate).[36]

[33] In the monitoring surveys, data collectors followed a standardized list of questions with categorical answer options (e.g., "professional," "unprofessional," or "a mess") to evaluate the condition of the vendors' equipment and hygiene practices. Each answer category was linked to a reference photo to ensure consistency across enumerators. These were the same photos shown to consumers in the consumer perception module, but they did not depict the specific items provided by the program. Questions regarding vendors' subjective evaluations of the equipment were included separately in Endline 1. The full list of questions and corresponding answer categories is provided in Tables A9 and A10 in the Appendix.

[34] To minimize the burden on vendors given the frequency of our visits, we structured the surveys as follows: during weekly monitoring in odd-numbered weeks, vendors responded to questions on business activities—including sales, expenditures, profits, and customer volume. In even-numbered weeks, they answered a different set of questions focused on food safety practices. During the post-treatment phase, when visits occurred every two weeks, both sets of questions (business practices and food safety norms) were included in each round. As noted, vendor interaction occurred only after the random audit had been completed.

[35] We use the "Hanna Instrument Free Chlorine Checker" (product link available here). For each sample, we record both the amount of free chlorine and the exact time and date of collection. Free chlorine—also known as residual chlorine—refers to the amount of chlorine available for disinfection.

[36] In the few instances where the vendor present during monitoring was not the kiosk owner, enumerators recorded that the person observed was not the owner. These cases were minimal.

4.5 Pre-Intervention Summary Statistics and Balance Checks

Columns (1) to (4) in Table 6 provides summary statistics and balance checks for a range of pre-intervention variables. We regress the left variable on strata fixed effects and report the parameter estimates for the treatments and the standard errors. On average, the kiosk owners are predominantly male, with an average age of 44 years, and 38% have at least primary-level education. Vendors typically have extensive experience and do not change kiosks frequently; the average vendor has been at their current kiosk for 19 years.[37] The kiosks themselves are small, with an average of 2 employees in total. Vendors work long hours, with kiosks open an average of 6 days a week, and stay at their stalls for around 13 hours per day. Most of this time is spent selling food (an average of 7.8 hours), 1.4 hours are spent on tasks defined as "cleaning," while the rest is dedicated to activities related to food preparation. Vendors report catering to an average of 76 consumers per day and earning an average of 660 ₹ in daily profits, which is approximately 8 USD.

In terms of food safety operations, as anticipated earlier, kiosks do have some of the equipment that we provide, although they are provisional and rudimentary, and few vendors possess or wear an apron or hairnets and display basic sanitary items such as a hand washing soap. Furthermore, only 18% of vendors in our sample have ever received any training on food safety practices and handling, and even less (5%) report treating their primary water source to make it potable. Finally, the last row of the table reports the levels of free chlorine in the primary water storage container. The average level is 0.16 per million (ppm), which is below the range 0.20-0.50, the minimum recommended level for drinking water recommended by the WHO.[38]

In terms of the balance between the assigned groups, joint orthogonality F-tests do not indicate any significant differences across the groups at the 10% level for the range of pre-treatment variables under consideration. Additionally, we conducted checks for balance between each treatment group and the control group, as well as among the different treatment groups separately. The results, in terms of p-values for difference-in-means tests, are presented in columns (5) to (8). Out of 80 coefficients, only 3 show some significance, whereas the rest are well balanced.

5 Estimation Strategy and Results

We now turn to estimating the causal effect of our intervention, which simultaneously reduces the cost of producing higher-quality (safer) meals and strengthens vendors' ability to signal improvements to consumers. First, we assess whether vendors utilize the provided equipment, whether we observe good hygiene practices during regular kiosk operations, and whether there are differential impacts between the two treatment groups. Second, we examine how sensitive vendors are to rising costs of maintaining quality by tracking outcomes after the end of the subsidy period. Third, we test whether vendors who use the large equipment are more likely to provide higher-quality meals (based on our pre-specified definition). Fourth, we evaluate the treatment effects on various business outcomes and labor supply measures. If consumer demand responds positively to quality upgrading, we expect to see increases in customer numbers, profits, and possibly prices among

[37] Nearly all vendors in our context are affiliated with unions, which are relatively informal organizations that serve as intermediaries between vendors and the municipality. Local leaders of these unions play a crucial role in managing the operations of the street food market in Kolkata.

[38] See technical note by WHO here. Similar recommendations are provided by the CDC here.

Table 6: Pre-Intervention Summary Statistics and Balancing

Variable	(1) Total Mean/(SE)	(2) Control Mean/(SE)	(3) Treatment 1 Mean/(SE)	(4) Treatment 2 Mean/(SE)	(5) Joint F-Test P-value	(6) (2)-(3) P-value	(7) (2)-(4) P-value	(8) (3)-(4) P-value
Demographics:								
Male	0.905	0.907	0.924	0.882	0.691	0.790	0.556	0.376
	(0.021)	(0.038)	(0.028)	(0.041)				
Age	43.938	44.175	44.033	43.565	0.923	0.893	0.819	0.709
	(0.760)	(1.437)	(1.355)	(1.117)				
At least primary education	0.376	0.423	0.326	0.376	0.243	0.105	0.421	0.317
	(0.032)	(0.052)	(0.063)	(0.049)				
Years at this kiosk	19.226	19.351	18.391	19.988	0.465	0.523	0.636	0.216
	(0.796)	(1.535)	(1.397)	(1.160)				
Business:								
Number of employees	2.288	2.412	2.152	2.294	0.629	0.276	0.697	0.587
	(0.126)	(0.187)	(0.243)	(0.235)				
Weekly number of work days	6.243	6.293	6.213	6.220	0.602	0.305	0.407	0.701
	(0.042)	(0.065)	(0.083)	(0.074)				
Hours of work per day	13.169	13.158	13.511	12.811	0.198	0.511	0.304	0.101
	(0.168)	(0.282)	(0.267)	(0.289)				
Hours spent selling food	7.806	7.799	8.129	7.463	0.118	0.427	0.209	0.052*
	(0.139)	(0.211)	(0.233)	(0.245)				
Hours spent cleaning	1.357	1.413	1.321	1.332	0.546	0.337	0.391	0.868
	(0.035)	(0.052)	(0.064)	(0.067)				
Daily number of customers	75.615	70.345	74.593	82.532	0.093*	0.546	0.018**	0.235
	(3.419)	(4.778)	(6.967)	(5.614)				
Daily profits	659.297	630.163	692.416	656.037	0.766	0.547	0.529	0.865
	(28.226)	(37.386)	(61.718)	(43.424)				
Food safety:								
Kiosk has handwashing facility	0.951	0.967	0.931	0.955	0.434	0.169	0.566	0.436
	(0.012)	(0.013)	(0.026)	(0.020)				
Kiosk has garbage bin	0.880	0.878	0.873	0.890	0.862	0.954	0.677	0.642
	(0.018)	(0.024)	(0.038)	(0.030)				
Kiosk has drinking water facility	0.996	0.994	1.000	0.994	0.331	0.306	0.973	0.197
	(0.003)	(0.005)	(0.000)	(0.006)				
Vendor uses apron	0.083	0.072	0.104	0.071	0.736	0.529	0.939	0.432
	(0.017)	(0.023)	(0.037)	(0.025)				
Vendor wears hair cover	0.012	0.011	0.017	0.006	0.730	0.582	0.655	0.416
	(0.005)	(0.007)	(0.013)	(0.007)				
Previous food safety training	0.179	0.175	0.174	0.188	0.992	0.953	0.898	0.935
	(0.031)	(0.043)	(0.038)	(0.076)				
Awareness [0,6]	2.226	2.186	2.261	2.235	0.896	0.611	0.778	0.941
	(0.097)	(0.131)	(0.140)	(0.232)				
Water:								
Treats main water source	0.045	0.072	0.040	0.019	0.265	0.627	0.132	0.306
	(0.016)	(0.035)	(0.026)	(0.012)				
Chlorine (ppm)	0.160	0.150	0.164	0.167	0.813	0.623	0.597	0.837
	(0.009)	(0.016)	(0.016)	(0.011)				

Notes: Data from the baseline and two pre-treatment monitoring surveys (surveys 1 and 2). Columns (1) to (4) present the parameter estimates and the associated standard errors for the treatment dummies, derived from a regression of the left variable on strata dummies and treatment dummies. Columns (5) to (8) report the p-values from the difference tests. Variables "male" through "number of employees" are from the baseline survey and have 284 observations; variables "work days" through "profits" are from the baseline and first pre-treatment monitoring surveys and have 543 observations; variables "handwashing facility" through "water treatment" are from the two pre-treatment monitoring surveys and have 523 observations. "Chlorine" is free chlorine parts per million, and has 698 observations. Panel (a) of Figure A10 in the Appendix illustrates the distribution of chlorine levels for the entire sample, and panel (b) for each area separately. The WHO recommends residual chlorine levels between 0.20 and 0.50 ppm for potable water (see technical note here). These visual representations show that, at baseline, the vast majority of vendors have a chlorine level in the water falling below the range of 0.20-0.50.

treatment group vendors. This may be sufficient to sustain higher food safety standards in the medium run.[39] Finally, we discuss underlying mechanisms that may explain the observed results.

5.1 Primary Outcome Variables

We create four primary outcome variables (indices) to measure quality: i) usage of high-quality large equipment; ii) usage of small supplies related to food safety; iii) compliance with food-safety practices; and iv) overall quality, which is the aggregate of the first three indices i)-iii). Our objective is to capture "best behavior," defined as the presence of equipment or practices considered "clean and professional."[40] Each outcome is expressed as a count of observed "best behaviors" within its respective category, ranging from zero to the maximum number of behaviors that can be observed in that category. Consequently, a higher count reflects better adherence to food safety standards. For example, if a vendor scores 3 out of 4 on a particular count variable, it indicates the vendor displayed three out of four possible instances of "best behavior" when monitored.[41]

First, the "large equipment" count variable ranges from zero to four and includes whether the kiosk displays a clean and professional-looking (i) hand-washing facility, (ii) primary water storage container, (iii) drinking water facility, and (iv) garbage bin.[42] Second, the "small supplies" count variable ranges from zero to three and includes whether the vendor is observed (i) wearing an apron, (ii) wearing a hair cover, and (iii) using the hand-washing facility with soap. Third, the "food-safety practices" variable ranges from zero to thirteen and includes practices such as (i) using soap to wash dishes, (ii) maintaining clean dishwater, (iii) keeping the garbage bin clean and empty, and avoiding (iv) visible garbage, (v) stagnant water, and (vi) food on the ground around the kiosk. Additional practices include (vii) using disposable plates, (viii) keeping the food preparation counter clean, (ix) using a clean towel, (x) covering cooked food, (xi) covering raw food, (xii) using utensils to handle food, and (xiii) washing hands before touching food. While the first two count indices capture the use of items supplied directly by the intervention, the food-safety practices index reflects unsubsidized, effort-intensive behaviours that vendors must carry out on their own. Together, these three component scores form our overall quality index, which ranges from 0 to 20 and captures the cumulative presence of observable sanitation-related inputs and behaviors.

In addition, we include the chlorine content in the kiosk's primary water source as a separate outcome. A binary variable equals one if chlorine levels exceed 0.20 ppm, indicating the water is safe to drink. We treat this separately from the small supplies index for two reasons: first, it is a binary rather than a count variable; second, chlorine levels were measured at different times from the main surveys and less frequently.

[39] All these hypotheses were pre-specified in our pre-analysis plan (PAP), registered with the AEA RCT Registry (AEARCTR-0008797).

[40] As mentioned earlier, enumerators were provided with illustrative pictures of what constitutes "clean and professional" behavior in this context.

[41] The construction of these variables follows the methodology outlined in our PAP. Although we pre-tested the questionnaire before each data collection round, some minor deviations occurred during the study period due to fieldwork challenges and efforts to improve data quality. These deviations are documented in Appendix D.

[42] The baseline large equipment measures in Table 6 are binary indicators of availability, ignoring quality. Consequently, treatment effect estimates for the large equipment are not directly comparable to the baseline means.

5.2 Estimation Approach

For the count outcome variables described earlier, we estimate intent-to-treat (ITT) effects using the following Poisson specification:[43]

$$\ln \mathrm{E}[Y_{i,c,t}|\mathbf{X}_{i,c,t}] = \beta_0 + \beta_1 T_{1,c,t} + \beta_2 T_{2,c,t} + \theta \bar{a}_{i,c,-1} + \mu_{strata} + \boldsymbol{W}'_{i,c,t}\boldsymbol{\gamma} \tag{1}$$

where $Y_{i,c,t}$ denotes the outcome for vendor i, in cluster c, at time t. The outcome vector includes the four indices introduced in the previous section: overall quality, large equipment usage, small supplies usage, and food-safety practices. The expected value of the outcome, $E[Y_{i,c,t}|\mathbf{X}_{i,c,t}] = e^{\mathbf{X}'_{i,c,t}\beta}$, is modeled as the exponential of a linear combination of explanatory variables. The model is estimated via maximum likelihood. For outcomes that are not counts, such as whether chlorine levels exceed 0.20ppm, or business and labor outcomes, we use OLS. We estimate a log-linear specification for the business and labor outcomes. The variables $T_{1,c,t}$ and $T_{2,c,t}$ are treatment dummies: $T_{1,c,t} = 1$ indicates assignment to the "equipment only" group, and $T_{2,c,t} = 1$ indicates assignment to the "equipment with training" group, with the control group as the reference category. The coefficients β_1 and β_2 capture the ITT effects of each treatment, and their difference reflects the marginal effect of training. $\bar{a}_{i,c,-1}$ denotes the pre-treatment average availability of sanitary equipment at the kiosk. μ_{strata} are strata fixed effects based on area and cluster size. The vector $\boldsymbol{W}_{i,c,t}$ includes a small set of control variables: survey wave fixed effects, interviewer fixed effects, pre-treatment number of kiosk employees, and pre-treatment vendor experience. Given that randomization occurs at the cluster level, we cluster standard errors accordingly. To adjust for multiple hypothesis testing (MHT), we report sharpened q-values alongside standard errors.

In our main specification, we restrict the analysis to the post-treatment period, pooling data from all monitoring rounds starting with round 5—when all treated vendors had received the intervention.[44] Finally, both in the main text and the appendix, all specifications of equation (1) include interactions between the treatment indicators and a dummy for the post-subsidy period—i.e., after the first endline survey in September 2022, when the free provision of small supplies and chlorine tablets ended. These interactions capture whether treated vendors experience a differential change in outcomes once the subsidy ends, over and above any time trends common to all vendors. This allows us to assess whether the treatment effects persist, attenuate, or disappear when vendors must bear the full costs of maintaining quality.

5.3 Treatment Effects on Equipment Usage and Practices

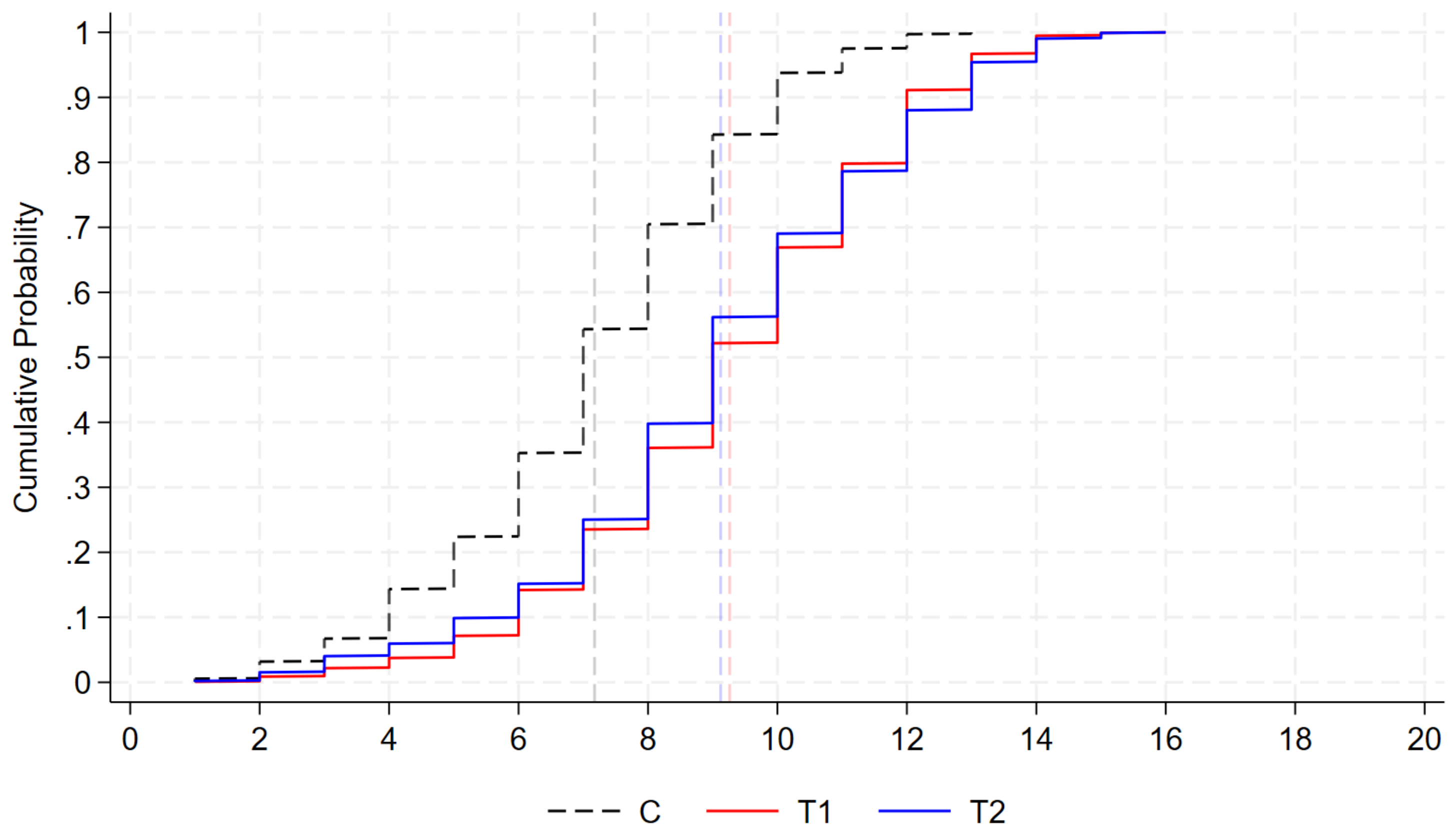
Figure 1 plots the cumulative distribution functions of our overall quality index, measured during the full observation period, including both the treatment and post-treatment phases. The figure shows a clear rightward shift in the distribution for both treatment groups compared to the control group, suggesting sustained differences in the adoption of quality-enhancing inputs and practices. However, there is little visual difference

[43] This Poisson specification was not pre-specified in the PAP. However, it follows standard practice in recent applied work with count outcomes which was published after our PAP (Chen and Roth, 2024; Wooldridge, 2023). As a robustness check, we also estimate OLS specifications as outlined in the PAP, and the results remain consistent.

[44] Pre-treatment controls are still included. This restriction ensures comparability across vendors and avoids issues from staggered implementation. As a robustness check, we also estimate an alternative specification that includes the two earlier post-treatment rounds, during which some vendors had not yet received the equipment. Results remain consistent.

between T1 and T2, suggesting minimal additional effect of the training.[45]

Figure 1: Distribution of Overall Quality by Treatment Group



Notes: Data from monitoring surveys (random audits) measured from period 5 (when the large equipment had been delivered to all vendors) up until the second endline survey in February 2023. Overall quality is defined as the total number of observed sanitation-related items and practices at the kiosk. The figure plots the empirical cumulative distribution functions of overall quality scores (range: 0–20). The vertical dashed lines represent group means.

Table 7 presents our main treatment effect estimates. Column (1) shows that treated vendors significantly improved their overall quality score: vendors in T1 and T2 increased their index by 0.30–0.26 units, respectively—amounting to increases of 34% and 30% relative to the control group mean of 7.4. Column (2) disaggregates quality by large equipment usage. Vendors in the control group used on average 0.95 items out of 4, while vendors in T1 and T2 used 1.09 and 0.98 more items, respectively—equivalent to a 199% and 166% increase. This corresponds to an average of 2.0–2.1 out of 4 large items in use (or roughly 50–53%). As shown in Table A13 in the Appendix, we observe large and significant gains for hand-washing stations (0.49–0.41), drinking water (0.37–0.26), and garbage bins (0.70–0.63). While these results indicate that many vendors used the equipment regularly, it is worth noting that full compliance was not achieved, despite the value of these facilities and the relatively low effort required for their use. We come back to this observation in the mechanisms section.

Column (3) considers the use of small supplies, with T1 and T2 vendors showing increases of 0.42 and 0.34 items (out of 3), corresponding to improvements of 40-52%. These smaller gains are driven largely by the use of aprons and hair covers (Table A13 in the Appendix). Column (4) focuses on food safety practices that were not directly subsidized. Treated vendors improved by about 0.10 points (10%). Benchmarked against comparable quality upgrading studies, this 10% improvement is meaningful but toward the lower end of reported effects (e.g., Bold et al., 2022; Hoffmann et al., 2023; Park et al., 2023; Deutschmann et al., 2023).

[45]Figure A11 in the Appendix presents cumulative distribution functions disaggregated by large and small items and food safety practices.

Table 7: Treatment Effects on Equipment Usage, Small Supplies and Food-Safety Practices

	(1) Overall quality (0-20)	(2) Large equip. (0-4)	(3) Small supplies (0-3)	(4) Food-safety practices (0-13)	(5) Chlorine $\mathbb{1}(>0.20)$ (ppm)
Equipment (T1)	0.295***	1.094***	0.416**	0.095**	0.887***
	(0.026)	(0.084)	(0.142)	(0.030)	(0.018)
	[0.001]	[0.001]	[0.004]	[0.003]	[0.001]
w/ training (T2)	0.259***	0.978***	0.335*	0.093**	0.903***
	(0.030)	(0.092)	(0.154)	(0.032)	(0.016)
	[0.001]	[0.001]	[0.023]	[0.005]	[0.001]
Equipment (T1) x post	-0.051	-0.177*	-0.239	-0.023	-0.841***
	(0.033)	(0.089)	(0.162)	(0.042)	(0.022)
	[0.063]	[0.031]	[0.063]	[0.172]	[0.001]
Equipment (T2) x post	-0.054	-0.149	-0.362*	-0.025	-0.849***
	(0.038)	(0.089)	(0.175)	(0.050)	(0.025)
	[0.063]	[0.054]	[0.028]	[0.172]	[0.001]
Control mean:	7.44	0.95	0.30	6.18	0.04
Implied T1 effect (%):	34.3	198.7	51.6	10.0	2144.6
Implied T2 effect (%):	29.6	165.9	39.8	9.7	2184.2
Clusters:	108	108	108	108	106
Observations:	3073	3073	3073	3073	2517
p-value T1-T2:	0.13	0.06	0.54	0.94	0.27

Notes: Data from monitoring surveys (random audits) measured from period 5 (when the large equipment had been delivered to all vendors) up until the second endline survey in February 2023. Outcome variables in Columns (1)-(4) are equal to the number of components of each count variable observed at the time of data collection. We use a Poisson regression model for estimation. Whereas, in Column (5) the outcome variable is a binary variable taking value one if the amount of chlorine is above 0.20 ppm, and zero otherwise. "post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. All regressions include strata fixed effects and the pre-treatment average sanitary equipment observed at the kiosk. To increase precision, we also include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Disaggregated results (Tables A14 and A15 in the Appendix) show improvements mainly in raw food coverage and garbage bin cleanliness, while other components, like towel/counter cleanliness or proper use of utensils, show no detectable change. Column (5) presents the largest treatment effects: the probability that chlorine levels exceed the 0.20ppm threshold increases by 89–90 percentage points in the treatment groups relative to the control mean of 4%. This implies near-universal compliance with water treatment among treated vendors. The last result confirms that providing chlorine tablets along with clear instructions is highly effective in overcoming barriers to adoption.[46]

Removing the subsidy. Next, we assess whether the treatment effects persist after the weekly provision of small supplies and chlorine tablets ends—when vendors must begin covering the ongoing cost of maintaining higher food safety standards themselves. If the effort or expense of procuring these inputs is sufficiently high, we might expect declines in both usage and overall quality. We examine this by looking at the interaction

[46] These results align with vendors' self-reported responses regarding whether they regularly use the large equipment and small supplies, find them useful, and perceive them as adding value to their business. Table A16 in the Appendix shows that vendors who received the equipment were much more likely to report regular usage of the large items, ranging from 87% to 97% across different components, than of the small supplies, for which reported usage ranges from 26% to 76%. While these self-reported figures likely overstate actual usage compared to our random audit data, they mirror the same pattern: larger equipment items are used more frequently than smaller ones. For each item, we observe a strong correlation between vendors' self-reported regular usage and their belief that the item adds value to their business and is appreciated by customers, supporting the behavioral mechanisms behind the observed effects.

term in our main specification. Because our specification includes monitoring period fixed effects, any general decline in equipment condition or cleanliness, are absorbed. The interaction terms therefore isolate the differential change in outcomes for treated vendors after the subsidy ends, relative to the control group.

Across treatment groups and outcomes in Table 7, all components of the quality index show negative interaction effects, and these declines are statistically significant (q-value < 0.10). The decrease in small supplies (Column 3) is unsurprising, as vendors were no longer receiving these items for free. The largest decline is in chlorine usage (Column 5): compliance drops from near-universal levels during the treatment phase to almost no vendors maintaining chlorine concentrations above the 0.20 ppm threshold in their primary water containers. This decline occurs despite clear guidance provided to treated vendors on where to purchase chlorine tablets locally.[47]

More puzzling is the observed decrease in large equipment usage (Column 2), as vendors received this equipment at no cost and faced no ongoing replacement expenses. A likely explanation is that using these items still requires regular attention and effort: containers need to be refilled, garbage bins emptied, and equipment surfaces kept clean.[48] We come back to this result in the mechanisms section. These findings suggest that even low-effort, durable improvements are difficult to sustain in complex, high-friction environments such as busy roadside kiosks, even after the initial investments have been made.

Food safety trainings. As shown in the bottom row of Table 7, we find no additional effects of food safety trainings.[49] The lack of increased adoption of safe food preparation practices may stem from vendors not perceiving the training sessions as particularly valuable. To explore this possibility, we collected data on vendors' perceptions of the training. Only around a quarter of T2 vendors reported that they found the training useful (26%), would recommend it to other vendors (26%), changed their behavior as a result (27%), or believed the training added value to their business (26%). When asked why they had not adopted the recommended practices, many cited unfamiliarity with the new behaviors or lack of time as the main barriers (Table A18 in the Appendix). These responses are consistent with data on vendor participation collected by our training team: while over 90% of vendors were rated as engaged or highly engaged during the first four weeks of the training, engagement dropped to 20% by week six and declined further to around 6% during the final four weeks (see Figure A13 in the Appendix).

Does equipment adoption drive better practices? A key question is whether the improvements we observe in unsubsidized food-safety practices (Column 4 of Table 7) were partly triggered by the visible infrastructure supplied through the intervention (Column 2 of Table 7). To test this, we conduct a mediation analysis following the potential outcomes framework of Imai et al. (2010).[50] We find that 78% of the treatment effect on

[47] Figure A12(a) in the Appendix shows the full distribution of chlorine levels, while panel (b) plots the share of vendors meeting the threshold over time. Both panels reveal a marked behavioral reversal once free distribution ceased.

[48] While we cannot empirically disentangle whether the decline in large item usage stems from depreciation, misuse, or reduced motivation to maintain the equipment, we provide some evidence in the mechanisms section suggesting that degradation and theft were contributing factors.

[49] In fact, T2 vendors even show statistically lower values for some equipment components (see Table A13 in the Appendix). Moreover, we observe relatively few differences between T1 and T2 even when focusing on the specific outcomes targeted by the training (see Tables A14 and A15 in the Appendix). In terms of self-reported behavior, Table A17 in the Appendix shows that T2 vendors report slightly better knowledge of hand-washing and more frequent water treatment, but also greater difficulty with hygiene tasks like emptying garbage. Most differences are small and not statistically significant, indicating limited changes in self-reported practices.

[50] We estimate mediation effects using the `mediate` command in Stata. Specifically, we assess whether the treatment effect on the food-safety practices index (0–13) operates through the adoption of large equipment (0–4). Both mediator and outcome are modeled with Poisson regressions. The specification

food-safety practices is mediated by the adoption of large equipment (Table A19 in the Appendix). Although this does not constitute a formal test of signaling, the magnitude of the mediated effect suggests that visible upgrades encouraged vendors to exert additional effort on cleanliness—either by reinforcing internal norms of professionalism or by helping vendors credibly signal higher standards to customers. These results are therefore consistent with the idea that infrastructure upgrades can indirectly promote effort-intensive hygienic behaviors.

Robustness checks and heterogeneous effects. We support these results with a battery of additional analyses. A more detailed explanation for each analysis is available in Appendix G. First, we replicate our main estimates including monitoring periods 3 and 4, during which the equipment was still being delivered. We find similar results (Table A20 in the Appendix). Second, we re-estimate the main specifications using OLS instead of Poisson, keeping the original count structure of the dependent variables. Results in Table A21 in the Appendix closely mirror those in Table 7. Third, we test for heterogeneity along three dimensions: geographic area, cluster size, and kiosk size. For area, we separate vendors by the three neighborhoods in Kolkata where the study was conducted. For cluster size, we rely on the natural urban clusters already described in the design section. For kiosk size, we classify vendors as operating either small kiosks (0–1 employees besides the owner) or large kiosks (2 or more employees). Overall, we find limited heterogeneity in treatment effects on the quality indices across areas (Tables A22-A25), cluster sizes (Tables A26-A29), and kiosk sizes (Tables A30-A33).[51] Lastly, Figure A14 in the Appendix plots the distributional treatment effects, indicating that the intervention was most effective in lifting vendors at the bottom of the quality distribution, though impacts on effort-intensive behaviors were limited.

5.4 Treatment Effects on Business Outcomes

Table 8 presents the impact of the treatment on various business outcomes (Panel A) and labor supply measures (Panel B). To increase statistical power, we pooled all observations into a single treatment group dummy. We also trim the top 2% of reported profits to mitigate the influence of potential outliers. Starting with Panel A, Column (1) shows that equipment provision led to a statistically significant 5.7% increase in monthly profits. This corresponds to an average increase of 3.1% relative to the value of the equipment provided (or approximately 36₹ more in profits per day).[52] This effect is on the lower end, but comparable to what the literature on returns to capital has found (e.g., De Mel et al., 2008; Jayachandran, 2021). This rise appears to be driven by increased turnover: vendors in the treatment group reported a 7.1% increase in customer volume (Column 4), roughly 5 additional customers per day, along with 6.4% and 7.6% increases in monthly sales and expenditures, respectively (Columns 2 and 3). Turning to Panel B, we find no evidence of treatment effects on labor supply, either at the extensive or intensive margin. Treated vendors did not extend their working time,

includes treatment assignment (T1 or T2), strata fixed effects, relevant pre-treatment covariates, and interactions between treatment and the mediator to allow for heterogeneous mediation. Standard errors are clustered at the vendor cluster level. The proportion mediated is calculated as the ratio of the estimated indirect effect (treatment $\rightarrow$ mediator $\rightarrow$ outcome) to the total effect. P-values are based on the delta method.

[51] We do find some differences in specific index components. For instance, treated vendors in Dalhousie and Sector V are significantly more likely than those in Hazra to have a professional-looking hand-washing station during the treatment period. These vendors are also more likely to maintain clean garbage bins and less likely to have garbage on the ground (Tables A23, A24, and A25 in the Appendix). However, these differences are not systematic and often reverse after the intervention ends.

[52] The equipment's total cost was 350 USD, or 29,117₹. A simple back-of-the-envelope calculation suggests that, with an average daily profit increase of 36₹, it would take approximately 809 days to recoup the investment (just over two years).

with no noticeable changes in the number of days worked per week or hours worked per day. Similarly, we find no significant differences in the time spent on preparation, selling, or cleaning.[53]

Table 8: Treatment Effects on Business Outcomes

	(1)	(2)	(3)	(4)	(5)
	Panel A: Business Outcomes				
	Profits, monthly	Sales, monthly	Expend., monthly	Custom., monthly	Prices
Equipment (T1 or T2)	0.057*	0.064*	0.076*	0.071*	-0.002
	(0.024)	(0.028)	(0.032)	(0.033)	(0.014)
	[0.081]	[0.081]	[0.081]	[0.081]	[0.563]
Control mean:	9.67	11.35	11.13	7.45	3.68
Implied T effect (%):	5.9	6.6	7.9	7.4	-0.2
Clusters:	108	108	108	108	103
Observations:	2761	2764	2764	2582	1565
Adjusted R^2:	0.52	0.70	0.68	0.55	0.98
	(1)	(2)	(3)	(4)	(5)
	Panel B: Labor Supply				
	Days, weekly	Total, daily hrs.	Prepare, daily hrs.	Sell, daily hrs.	Cleaning daily hrs.
Equipment (T1 or T2)	0.002	0.011	0.018	0.002	0.019
	(0.006)	(0.010)	(0.012)	(0.014)	(0.014)
	[0.563]	[0.297]	[0.176]	[0.563]	[0.250]
Control mean:	1.80	2.51	1.30	1.95	0.30
Implied T effect (%):	0.2	1.1	1.9	0.2	1.9
Clusters:	108	108	108	108	108
Observations:	2764	2769	2769	2769	2769
Adjusted R^2:	0.18	0.49	0.24	0.51	0.49

Notes: Data from monitoring surveys (random audits), except for price data, which comes exclusively from Endline 2. All outcome variables are expressed in logs. Profits, sales, expenditures, and prices are reported in rupees. "Profits, monthly", "Sales, monthly," "Expenditures, monthly," and "Customers, monthly" are computed by multiplying daily values of the variable by the number of days worked in the previous week, then multiplying by four. “Prices” refers to the price of each item sold at a kiosk. “Days, weekly” refers to the number of days the kiosk was open in the previous week. “Total, daily hrs.” refers to the average number of hours per day the kiosk was open. “Prepare, daily hrs.” refers to the average number of hours per day the vendor spent preparing food. “Sell, daily hrs.” refers to average daily hours spent selling. “Cleaning, daily hrs.” refers to the average number of hours per day the vendor spent cleaning the kiosk. All OLS regressions, except for prices, are conducted at the vendor level and pooled across all periods from period 5 (when the large equipment had been delivered to all vendors) up to the second endline survey in February 2023. We also trim the top 2% of reported profits to mitigate the influence of potential outliers. Whereas, the price regression in Column (5) is conducted at the item level and includes controls for baseline prices. “Equipment (T1 or T2)” is an indicator equal to one if the vendor belongs to treatment groups T1 or T2, and zero otherwise. All regressions include strata fixed effects and the logged pre-treatment average of the outcome measured at baseline. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period and interviewer, as well as controls for number of employees, years of experience, and whether the vendor keeps accounting records. Standard errors are clustered at the vendor cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Regarding prices, we report a precise null effect in Column (5) of Panel A, which is particularly notable given that (i) the value of the equipment transfer was substantial, and (ii) street food consumers report a strong willingness to pay higher prices for food served from cleaner kiosks by cleaner vendors. To better understand this result, we included a set of Likert-type questions in our second endline survey in February 2023, asking vendors about the local institutional environment. For the price items, to maintain neutral wording, we asked

[53] We also find no differences in menu selection, expected days of work, or expected profits over the next 7 days. In addition, there is no effect on total business assets, the likelihood of reporting any savings, or the amount saved per month (see Table A34 in the Appendix).

about "different prices" rather than "higher" or "lower" prices. Results are reported in Figure A15 in the Appendix. Interestingly, we find evidence of local coordination in both pricing and menu choices. Two third of vendors acknowledged some degree of price coordination, and 86% agreed that substantial deviations from prevailing prices could trigger social repercussions within their community. Similarly, 72% of vendors reported coordination in menu choices, and 87% believed that offering highly similar meals to other vendors could result in negative social consequences.[54]

We further check for evidence of price homogeneity in Figure A16 in the Appendix. We focus on *thali*, a relatively homogeneous and widely sold item, stratified by protein type. Overall, the four types of thali represent 26% of all items sold by vendors in our sample. Panel (a) demonstrates that price variation is driven predominantly by differences between clusters, with less than 10% attributed to variation within clusters. This is corroborated by Panel (b), which shows that prices vary by only 5–7% on average within clusters. Altogether, these findings suggest that vendors operate in a tightly coordinated market with limited scope for strategic differentiation. As a result, those seeking to increase profits by upgrading kiosk quality are unable to do so by raising prices and must instead rely on attracting more customers, making customer volume, rather than price, the primary channel for profit growth.

These patterns align with evidence that anti-competitive norms restrict strategic deviation. For example, in a related setting, Banerjee et al. (2024) document that sanctions deter price cuts and that prices in treated markets track controls even as sales rise. Our setting differs because a quality upgrade could justify higher prices, yet the same norm logic can suppress any price move: upward deviations may be viewed as "greedy," violating a focal community price and triggering social penalties. A simple focal-price norm can rationalize flat prices with higher customer volume and the positional spillovers that we document in the next section. The limitation, of course, is that our evidence on norms is self-reported, prices are observed once at baseline and endline, and punishments are not directly documented; testing asymmetric sanctions and experimentally shifting focal prices would address these gaps.

Does quality drive business gains? Similar to our analysis of hygiene practices, we use mediation analysis to assess whether the observed business improvements operate through enhanced vendor quality. Specifically, we examine whether overall quality mediates the treatment's effect on profits, sales, expenditures, and customer numbers. As shown in Table A19 in the Appendix, we find that 41% of the effect on customer volume and roughly 20–23% of the effects on other business outcomes are mediated by overall quality. While mediation estimates do not establish a causal signaling mechanism in the formal sense, the magnitude and consistency of the results suggest that visible improvements in cleanliness and infrastructure likely contributed to stronger customer engagement. These findings, together with the stated preference data indicating a high willingness to pay for clean food environments, are consistent with the interpretation that infrastructure upgrades enhanced vendors' perceived credibility and helped attract more customers—possibly by making quality more salient or easier to assess.

[54]Despite the absence of formal "menu costs," price revisions are infrequent: nearly 80% of vendors update prices only once per year, and another 10% do so even less frequently. Vendors also report high awareness of competitors' offerings: 95% know what meals others sell, and 80% know their prices. Nearly 95% attribute fluctuations in profits primarily to competition from other street food vendors, rather than from formal food establishments. These patterns reinforce the idea that price adjustments are constrained by local norms and social pressures.

Robustness checks. We conduct several robustness checks to verify the consistency of our results. First, we repeat the estimation of our main specification by including observations from periods 3 and 4, when equipment was being delivered, and find broadly similar results (Table A35 in the Appendix). Second, we replicate the analysis without trimming the top 2% of profit values. The estimated effects remain of similar magnitude, although significance is somewhat reduced for profits and customer numbers, while the effects on sales and expenditures stay robust (Table A36 in the Appendix). Third, we re-estimate the models using daily outcomes (as reported by vendors during data collection) rather than monthly aggregates. Point estimates are slightly smaller for some outcomes, but statistical significance improves overall (Table A37 in the Appendix). Fourth, we estimate the treatment effects in levels rather than logs. In this specification, the profit effect becomes statistically insignificant, though the effects on sales and expenditures remain sizeable and significant (Table A38 in the Appendix). Finally, to improve estimation precision, we apply Post-Double Selection Lasso (Belloni et al., 2013) to select relevant baseline covariates that are most predictive of each outcome. Results on profits are identical to the main results, and those on sales, expenditure and customers slightly weaker (Table A39 in the Appendix). Taken together, while the specific estimates vary slightly across specifications, the overall pattern consistently points to improvements in business performance through higher sales.

5.5 Spillover effects

One of the core assumptions of our experimental design is that there are no spillovers across clusters of vendors. This assumption is plausible when focusing on vendor-to-vendor interactions: vendors in different clusters are typically separated by at least 30 meters, limiting their ability to observe each other's equipment, practices, or training. However, the same may not hold for consumers, who are more mobile and may shop across clusters. If consumers observe improvements in a treated cluster—such as cleaner equipment or better hygiene—they may shift their purchasing behavior toward those vendors. In that case, observed improvements in quality or business outcomes may partly reflect a reallocation of demand across vendors, rather than an overall increase in quality or market size. To investigate whether our results are driven by such consumer-side externalities, we estimate a spatial spillover model that tests whether the treatment indirectly affects nearby untreated vendors by altering the competitive landscape. We follow the approach of Miguel and Kremer (2004),[55] augmenting it with spatial vendor counts at different radii. For simplicity, we aggregate T1 and T2 into a single treatment group:

$$\begin{aligned} \ln \mathrm{E}[Y_{i,c,t}|\mathbf{X}_{i,c,t}] = \beta_0 + \beta_1 T_{c,t} + \sum_d (\gamma_d \cdot [N^T_{d,i} - \bar{N}^T_{d,i}]) + \sum_d (\phi_d \cdot [N_{d,i} - \bar{N}_{d,i}]) + \\ + \theta \bar{a}_{i,c,-1} + \mu_{strata} + \boldsymbol{W}'_{i,c,t}\boldsymbol{\gamma} \end{aligned} \tag{2}$$

Here, $N^T_{d,i}$ indicates the number of treated vendors within distance d of vendor i, while $N_{d,i}$ denotes the total number of vendors (treated or not) in the same radius. To focus on local deviations, we subtract the sample mean within each radius. Each vendor has on average 37 other vendors within 400 meters, of whom 24 are treated (i.e., assigned to either T1 or T2). We define two bands of proximity: "close" vendors within 0–400 meters (roughly a 5-minute walk), and "distant" vendors between 400 and 800 meters (up to 10 minutes). This

[55] The literature offers several approaches to estimating spillover effects, including alternative spatial and network-based models. We adopt this specification because it was pre-specified in our PAP. In slight abuse of notation, the estimating equation (2) is written following the specification of the Poisson model. We use this specification for the quality indices. Whereas, for business outcomes we adopt as before a log-lin specification.

specification mirrors the structure used in the consumer DCE.

Table 9: Treatment Effects on Quality and Business Outcomes with Spillover Effects

	(1) Overall quality (0-20)	(2) Large equip. (0-4)	(3) Profits, monthly	(4) Sales, monthly	(5) Custom., monthly
Equipment (T1 or T2)	0.246***	0.917***	0.056*	0.072*	0.064*
	(0.023)	(0.067)	(0.023)	(0.028)	(0.032)
	[0.040]	[0.040]	[0.040]	[0.040]	[0.058]
Vendors 0-400m	0.005	0.027***	0.007**	0.005	0.004
	(0.003)	(0.008)	(0.002)	(0.003)	(0.003)
	[0.037]	[0.093]	[0.037]	[0.093]	[0.118]
Vendors 400-800m	0.004	0.034***	0.009*	0.006	0.006
	(0.003)	(0.008)	(0.003)	(0.004)	(0.005)
	[0.040]	[0.102]	[0.040]	[0.102]	[0.116]
T1 or T2 vendors 0-400m	-0.007	-0.039**	-0.013**	-0.011*	-0.009
	(0.005)	(0.012)	(0.004)	(0.005)	(0.005)
	[0.022]	[0.046]	[0.022]	[0.046]	[0.064]
T1 or T2 vendors 400-800m	-0.007	-0.052***	-0.013*	-0.009	-0.011
	(0.005)	(0.013)	(0.006)	(0.007)	(0.008)
	[0.046]	[0.116]	[0.046]	[0.116]	[0.116]
Control mean:	7.44	0.95	9.69	11.36	7.46
Clusters:	108	108	108	108	108
Observations:	3073	3073	2845	2848	2650

Notes: Data from monitoring surveys (random audits) measured from period 5 (when the large equipment had been delivered to all vendors) up until the second endline survey in February 2023. Outcome variable definitions and estimation details can be found in Tables 7 and 8. We aggregate treatment groups for convenience. "Vendors 0-400m" is calculated as the number of vendors within a 0 to 400m radius centered around the mean; "Vendors 400-800m" is defined similarly. All regressions include strata fixed effects and the pre-treatment average sanitary equipment observed at the kiosk. To increase precision, we also include fixed effects for the survey period, interviewer, number of employees, and years of experience. Standard errors at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table 9 reports spillover estimates for selected outcomes.[56] The central result is negative spillovers on business outcomes: as the number of nearby treated vendors rises, untreated vendors record lower profits, sales, and customer volume (Columns 3–5). These effects are statistically significant in both the 0–400 m and 400–800 m bands (q-value < 0.10).[57] These findings suggest that the treatment's success in improving business outcomes is partly driven by relative visibility: treated vendors benefit most when they are one of few in their local area to receive support, making their higher standards more noticeable to consumers. However, as more nearby vendors are treated, this signaling advantage diminishes, and the returns to quality decline. In other words, treatment effects are positional: they help vendors stand out and capture a larger share of existing customers, rather than expanding the total customer base.

A final remark concerns the competitive nature of returns to upgrading. Our combined results indicate that local street food revenues are essentially zero-sum. This pattern points to reallocation rather than market expansion, implying a transfer in surplus from (some) producers to consumers and from nearby competitors to

[56] The results for the remaining outcomes are in Section I.

[57] In contrast, we find no evidence of spillovers for less visible outcomes, such as small supplies or food practices (Table A40 in the Appendix). We see some negative spillover effects for labor supply responses, in particular days and hours worked (Table A41 in the Appendix). Results are robust to removing the mean-centering (Table A42 in the Appendix), but the effects attenuate when using inverse-distance weighting (Tables A43 and A44 in the Appendix).

treated firms. In such an environment, coordinated vendors (like in our case) have weak incentives to adopt if adoption merely redistributes customers without allowing higher markups. Moreover, the private return to upgrading is positional and declines as more neighbors upgrade. Quantitatively, our profits specification shows an own-treatment effect 0.056 log points, while each additional upgraded neighbor within 0–400 m reduces own profits by 0.013 log points. Holding local vendor density at its mean (37 neighbors, 24 treated), the "break-even point" occurs once the treated share rises about 12 percentage points above the mean (roughly from 65% to 77%, which is 4–5 additional upgraded neighbors). Beyond this threshold, the private profit return to upgrading becomes negligible. Taken together, these patterns suggest that, absent enforceable standards or coordinated incentives, widespread voluntary upgrading is unlikely to sustain itself.

5.6 Potential Mechanisms

Recall that our intervention was built on a simple premise: if vendors could upgrade at low cost and if customers could recognize those upgrades, then maintaining cleanliness should become a profitable and self-sustaining strategy. In the previous sections, we showed that the program led to better food safety practices and higher profits, indicating that it worked in the intended direction. However, the size of these effects was modest. Improvements in quality—particularly when focusing on effort-dependent behaviors—fall at the lower end of what the quality-upgrading literature has documented. Profit gains, while comparable to returns seen in other capital-subsidy programs, remain small in practical terms. Importantly, vendors appear to be gradually returning to their original low-quality equilibrium. This raises a central question: why is it still so difficult for vendors to sustain higher food safety standards, even after a large, direct injection of capital and support into the market? Given the logic of the intervention, the answer likely lies in one of two factors: either the effort or cost of maintaining higher standards is still too great, or the improvements are not visible enough to consumers to generate meaningful business returns. We designed our second endline survey to probe more deeply into why vendors struggle to maintain higher food safety standards. Results are documented below.

Local environmental constraints. The most striking findings come from vendors' own accounts of the physical and structural barriers they face. Results are reported in Figure A15 in the Appendix. First, theft emerges as a key concern: kiosks are semi-permanent and exposed, with limited options for securing valuable equipment overnight, and vendors cannot legally insure their tools given their informal status. Nearly all vendors (99%) agreed that theft of expensive equipment is common in their area, and 95% cited it as a major barrier to investing in large items. These risks translate into real costs (e.g., 15 vendors (9%) reported theft of the equipment we provided during the study), and lead many to invest extra effort in protecting the equipment, with some removing items from their kiosks entirely. Among those who did not regularly use the large items provided, nearly all reported taking them home for personal use, which protects the equipment but adds the daily burden of transport and setup. This appears to be the most likely explanation for why compliance with the use of large equipment was 50–53%.

Second, vendors operate in highly unsanitary and challenging environments—dusty roads, stagnant water, open garbage, and heavy foot traffic make it difficult to keep equipment clean. In these conditions, items can quickly become dirty, damaged, or worn, and may require basic maintenance to remain functional. Vendors often lack the resources or incentives to repair or replace broken components, which may help explain the observed declines in usage over time. Even vendors who regularly used the provided equipment and were initially observed to have "clean and professional" setups showed visible deterioration just three months later. Together, these factors show that maintaining cleanliness comes with substantial and recurring costs: securing equipment and preserving hygiene in adverse conditions demands substantial daily effort that cannot be subsidized.

Limited consumer visibility. Another plausible explanation for the modest profit gains is limited consumer visibility of food-safety improvements. We could not collect additional consumer data at endline to test this mechanism directly, but baseline evidence showed that consumers struggle to detect unsafe practices. If customers cannot reliably recognize improvements—even when vendors install cleaner or newer equipment—upgrades may fail to translate into stronger or immediate demand responses, particularly in a market where most kiosks look superficially similar. As a result, the business payoff from maintaining higher standards may remain too small or uncertain to offset the daily effort required.

Other potential mechanisms. While the environmental and visibility constraints discussed above likely explain much of the observed modest effect and decline in quality, we cannot rule out the influence of other factors. First, vendors may be deliberately reducing their efforts after gaining more customers. If customers do not quickly notice or react to a drop in hygiene, vendors might continue to enjoy higher sales even as they gradually stop using equipment or following food safety practices. Although our sample is not large enough to test this channel directly, time-series patterns offer little support for this idea. As shown in Figure A17 in the Appendix, the profit difference between treated and control vendors seems to decrease slightly after the end of the treatment period, even though the number of customers stays relatively constant. This suggests that vendors are not increasing their earnings by cutting corners. In addition, we find no differences in reported cleaning hours between treatment and control groups, implying that hygiene declines are not due to reduced effort on daily cleaning tasks. Second, there may be structural barriers that prevent vendors from turning quality improvements into higher profits. Our baseline data show that most vendors have worked in the same fixed spot for nearly 20 years, with little space to expand due to sidewalk congestion and informal zoning restrictions. Since prices are rigid and competition is not based on price, the only way to earn more would be to attract more customers or work longer hours. However, as shown in Table 8, vendors do not increase their labor supply in response to treatment, likely because they already work long hours (13 per day, six days a week on average), leaving little room to do more. Hence, even when vendors improve food safety, their ability to turn those improvements into lasting financial gains may be limited by external constraints and the realities of informal street vending.

6 Conceptual Framework

To help interpret the empirical findings, we outline a simple model of vendor decision-making under moral hazard with imperfect monitoring. The goal is to rationalize the key dynamics observed in the field: partial and temporary compliance with food safety standards, modest business gains, and post-intervention reversals. The framework rests on four features of the setting: (i) vendors possess some local market power and serve repeat customers; (ii) quality is binary and not perfectly observable;[58] (iii) consumers value cleanliness; and (iv) the intervention both lowers the cost of producing high quality and signals improved standards to consumers.

6.1 Vendor Decision Problem

Suppose each vendor can produce either low-quality (L) or high-quality (H) food. High quality requires (a) a one-time investment in large, highly visible equipment (fixed cost F) and (b) a per-unit cost of safe practices $c_s > c_0$, where c_0 is the baseline unit cost of producing low quality, and c_s is the unit cost to maintain hygienic practices. Low quality yields a constant per-unit profit of $(p-c_0)$, where p is exogenous to the quality decision.[59] Demand under low quality is denoted $D_L(p)$, with $\partial D_L/\partial p < 0$. In the absence of intervention, vendors choose L because upgrading requires purchasing large equipment and incurring a higher per-unit cost, both of which are prohibitively expensive. Hence, baseline profits are:

$$\pi(p) = (p-c_0)D_L(p).$$

The intervention has two effects. First, it removes or subsidizes the fixed cost F and temporarily covers c_s, making high-quality production financially viable. Second, it introduces a visible signal—sanitation equipment—that allows treated vendors to access the higher demand curve $D_H(p) > D_L(p)$, reflecting consumer willingness to pay for cleanliness. Once equipped, a treated vendor may either (a) truly maintain high standards (H) or (b) shirk on safe practices and supply low quality (L) while still displaying the equipment.

If a treated vendor chooses to produce H, they earn:

$$\pi_H^T(p) = (p-c_s)D_H(p).$$

Alternatively, if they shirk, using the equipment signal but not adopting hygienic practices, they earn:

$$\pi_L^T(p) = (p-c_0)(1-\lambda)D_H(p),$$

where $\lambda \in [0,1]$ captures the fraction of consumers who detect shirking, either through direct observation, post-purchase experience, or word-of-mouth.[60,61,62]

---

[58] That is, consumers can observe visible equipment but cannot monitor ongoing maintenance. This distinction is central to our modeling approach and reflects the idea that consistent food safety requires two key investments: costly equipment and continuous, "behind-the-scenes" efforts in safe food handling practices.

[59] As shown in Section 5, prices remain virtually identical across kiosks and change at most once a year.

[60] In this simplified conceptual framework, we model detection as a fraction λ of consumers who *automatically* identify low-effort vendors. In a more general framework, each consumer would observe a noisy signal, drawn from different distributions depending on whether the vendor is genuinely high or low quality. Formally, if $s \sim F(\cdot \mid H)$ when the vendor exerts high effort and $s \sim F(\cdot \mid L)$ otherwise, a consumer updates beliefs via Bayes' rule and buys at price p only if the likelihood ratio $\ell(s) = \frac{f(s|H)}{f(s|L)}$ exceeds a threshold ℓ^*. Defining $\Phi_H = \Pr\big(\ell(s) > \ell^* \mid H\big)$ and $\Phi_L = \Pr\big(\ell(s) > \ell^* \mid L\big)$, the profit expressions become $\pi_H^T(p) = [p-c_s]\Phi_H D_H(p)$ and $\pi_L^T(p) = [p-c_0]\Phi_L D_H(p)$. Truthful high-quality production is sustained when $[p-c_s]\Phi_H > [p-c_0]\Phi_L$, which nests the simpler condition $p\lambda + c_0(1-\lambda) > c_s$ in equation (3) when we set $\Phi_H = 1$ and $\Phi_L = 1-\lambda$. See Appendix J for a fuller discussion of the noisy-signal model and its implications.

[61] Our consumer data suggests that detection is plausible. Indeed, 41% of respondents report finding it easy to identify unsafe food, and 54% say they would switch vendors after experiencing food-related illness. We acknowledge that the "41% detect unsafe food" figure offers only partial evidence about detection probabilities, not a direct measure of λ. Consumers' detection power in more realistic signal-detection models (e.g., Appendix J) is likely heterogeneous and depends on both Type I and Type II errors. For our purposes, however, this statistic is sufficient to suggest that $\lambda > 0$.

[62] While λ denotes static detection probability, it may also capture reputational forces. Vendors often serve repeat customers who are better positioned to detect hygiene lapses over time. These regulars may stop purchasing if quality falls, generating sustained demand losses. Thus, λ reflects both immediate detection and longer-run reputational risks. This interpretation aligns with dynamic reputation models such as Bai (2024), though a full dynamic treatment

6.2 Equilibrium Behavior

The key condition for sustained high-quality production among treated vendors is that the profit from truthful behavior exceeds the profit from shirking:

$$\pi_H^T(p) > \pi_L^T(p).$$

Substituting the expressions above and simplifying, we obtain:

$$p\lambda + c_0(1-\lambda) > c_s. \tag{3}$$

This inequality captures the key incentive condition: high-quality production is sustainable only when the cost of maintaining safe practices is low *and* the risk of detection is sufficiently high. If either condition worsens—costs rise or monitoring weakens—vendors will no longer find it profitable to maintain high standards.

6.3 Mapping to Empirical Findings

The dynamics in equation (3) provide a simple lens through which to interpret our main results. At the start of the intervention, both F and c_s are close to zero due to the provision of large equipment and free weekly supplies. The detection parameter λ is unaffected by the intervention; however, even with a modest detection rate, the inequality is likely to hold when costs are low, leading to improved vendor behavior. Consistent with this prediction, Section 5.3 documents a substantial increase in equipment usage and safe practices (Columns 2–4 of Table 7), along with a sharp rise in potable water usage (Column 5).

Once the subsidy ends, however, c_s rises significantly. Vendors must now purchase chlorine, replace aprons, and invest daily effort to maintain cleanliness. Although the equipment remains available, vendors face the full marginal cost of hygienic practices. According to condition (3), unless λ or p increase sufficiently to offset these costs, the incentive to maintain quality weakens. Empirically, this is precisely what we observe: equipment use declines, chlorine use collapses, and hygiene practices stagnate – reflected in the negative interaction coefficients in Table 7.

Finally, with p fixed, improved quality increases profits only by boosting $D_H(p)$. This limits vendors' ability to recover rising costs, especially in a crowded market. Section 5 finds a 5–7% increase in profits—small, but consistent with limited horizontal demand shifts in a dense competitive environment. Meanwhile, the detection rate λ is likely moderate: although many consumers report valuing cleanliness, few have the understanding to recognize conditions conducive to lower risk of food poisoning. Thus, the right-hand side of condition (3) increases, while the left-hand side does not.

7 Conclusion

Street food is an important source of nutrition for urban-dwellers in low- and middle-income countries; however, it is also frequently considered a public health risk due to high levels of food contamination. In this paper,

lies beyond the scope of this paper.

we document strong consumer demand for safer food and design an intervention enabling vendors to upgrade their kiosks in ways that are visible to consumers, thereby incentivizing safer practices through profitability. While we observe improvements in food safety and increases in profits, these gains are limited and appear to dissipate over time. This raises the question of whether such interventions can generate lasting improvements. Higher customer numbers and increased profits among treated vendors suggest that improvements are feasible. However, the precarious nature of their businesses—exposed to extortion and theft, and with limited capacity to raise prices or scale operations—makes sustained change difficult. The actual returns to equipment are likely lower than anticipated, highlighting the challenges of addressing moral hazard in informal, "survivalist" enterprises.

Our findings offer policy insights tailored to local conditions. Vendors generally use equipment when it is provided, but are unlikely to purchase it themselves, and large-scale public provision of high-cost items would be prohibitively expensive. A more feasible approach may involve distributing low-cost, high-impact items—such as chlorine tablets and soap—free of charge. Given the importance of soap and potable water for safe food preparation, such policies may offer higher returns on investment. The use of hygiene inspectors to coerce vendors to provide higher quality is another avenue; however, without accompanying price increases, this would likely be quite burdensome for vendors. It would also require sustained commitment from municipalities to allocate resources for consistent oversight.

The broader regulatory environment regarding street vending is clearly a critical component for food-safety improvements. India's 2014 Street Vending Act addresses some key barriers, but many provisions remain unimplemented. Local discussions are underway to begin licensing vendors, but formal conditions for issuing licenses have yet to be defined.[63] To our knowledge, food safety standards are not currently part of these discussions, aside from vague directives to keep areas clean.

Our results also indicate a greater need to protect vendors from theft and extortion, which would encourage greater investment in hygienic improvements. Many evictions and bribes stem from confusion over complex legal frameworks and the absence of clear enforcement channels. Simplifying licensing could reduce harassment and create stronger incentives for safe practices. Municipal services are equally critical: access to clean water, sanitation points, waste collection, and a reliable mechanism to report abuse would help sustain improvements. The creation of dedicated vending zones—another feature of the 2014 Act—could support better hygiene if they guarantee sanitation and service provision. However, implementation faces practical challenges. In Kolkata, for instance, space constraints in central areas limit such efforts, and a designated vending area in Sector V remains unused due to concerns over low foot traffic. In the short to medium term, policy efforts would be better directed at upgrading hygiene where vendors currently operate. Small, low-cost investments in sanitation hardware, clearer rules, and regular training can build trust and visible improvements. Over time, coordinated legal, infrastructural, and service reforms could provide a realistic path to safer, more sustainable street food systems.

[63] The state government of West Bengal has directed that kiosks must i) be on pavement, not road; ii) take up a third or less of the width of pavement, leaving the rest for pedestrians; iii) operate at a minimum distance from major crossings and store entrances, and iv) be free of tarpaulin and other inflammable material.

References

Abrahale, K, Sandra Sousa, Gabriela Albuquerque, Patrícia Padrão, and Nuno Lunet (2019) "Street food research worldwide: A scoping review," *Journal of Human Nutrition and Dietetics*, 32 (2), 152–174.

Ambler, Kate, Mike Murphy, Sinh Dang-Xuan et al. (2024) "Business and public health impacts of a food safety rating program among pork vendors in Vietnam," IFPRI Working Paper December 2024, Washington, DC.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja (2017) "Report cards: The impact of providing school and child test scores on educational markets," *American Economic Review*, 107 (6), 1535–1563.

Andrade, Anderson Assunção, Aline Dias Paiva, and Alessandra Barbosa Ferreira Machado (2023) "Microbiology of street food: Understanding risks to improve safety," *Journal of Applied Microbiology*, 134 (8), lxad167.

Arnold, Benjamin F and John M Colford (2007) "Treating water with chlorine at point-of-use to improve water quality and reduce child diarrhea in developing countries: A systematic review and meta-analysis," *The American Journal of Tropical Medicine and Hygiene*, 76 (2), 354–364.

Ashraf, Nava, James Berry, and Jesse M Shapiro (2010) "Can higher prices stimulate product use? Evidence from a field experiment in Zambia," *American Economic Review*, 100 (5), 2383–2413.

Bai, Jie (2024) "Melons as lemons: Asymmetric information, consumer learning and seller reputation," *The Review of Economic Studies*, Forthcoming.

Banerjee, Abhijit V., Greg Fischer, Dean Karlan, Matt Lowe, and Benjamin N. Roth (2024) "Do Collusive Norms Maximize Profits? Evidence from a Vegetable Market Experiment in India," Working Paper 23-006, Harvard Business School, https://www.hbs.edu/ris/Publication%20Files/23-006_a419a5d5-b918-430f-bb1e-0d6a86afd77e.pdf.

Belloni, Alexandre, Victor Chernozhukov et al. (2013) "Least Squares After Model Selection in High-dimensional Sparse Models," *Bernoulli*, 19 (2), 521–547.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli (2006) "Adaptive linear step-up procedures that control the false discovery rate," *Biometrika*, 93 (3), 491–507.

Bennett, Daniel and Wesley Yin (2019) "The market for high-quality medicine: Retail chain entry and drug quality in India," *The Review of Economics and Statistics*, 101 (1), 76–90.

Bivins, Aaron, Sarah Lowry, Sonal Wankhede, Rajashree Hajare, Heather M Murphy, Mark Borchardt, Pawan Labhasetwar, and Joe Brown (2021) "Microbial water quality improvement associated with transitioning from intermittent to continuous water supply in Nagpur, India," *Water Research*, 201, 117301.

Björkman Nyqvist, Martina, Jakob Svensson, and David Yanagizawa-Drott (2021) "Can good products drive out bad? A randomized intervention in the antimalarial medicine market in Uganda," *Journal of the European Economic Association*, 20 (3), 957–1000.

Blair, Graeme, Jasper Cooper, Alexander Coppock, and Macartan Humphreys (2019) "Declaring and diagnosing research designs," *American Political Science Review*, 113 (3), 838–859.

Blum, Annalise G, Clair Null, and Vivian Hoffmann (2014) "Marketing household water treatment: Willingness to pay results from an experiment in rural Kenya," *Water*, 6 (7), 1873–1886.

Bold, Tessa, Selene Ghisolfi, Frances Nsonzi, and Jakob Svensson (2022) "Market access and quality upgrading: Evidence from four field experiments," *American Economic Review*, 112 (8), 2518–52.

Breza, Emily, Supreet Kaur, and Nandita Krishnaswamy (2019) "Scabs: The Social Suppression of Labor Supply," NBER Working Paper 25880, National Bureau of Economic Research, 10.3386/w25880.

Central Ground Water Board (2022-2023) "Ground Water Quality of West Bengal," Report.

Chen, Daniel and Christopher Roth (2024) "What makes a job good? Evidence from job search in Uganda," *The Quarterly Journal of Economics*, 139 (2), 891–946.

Choudhury, Manisha, Lipi Mahanta, Jayashree Goswami, Minakshi Mazumder, and Barnali Pegoo (2011) "Socio-economic profile and food safety knowledge and practice of street food vendors in the city of Guwahati, Assam, India," *Food Control*, 22 (2), 196–203.

Chukuezi, Comfort O (2010) "Food safety and hygienic practices of street food vendors in Owerri, Nigeria," *Studies in Sociology of Science*, 1 (1), 50–57.

Clasen, Thomas, Wolf-Peter Schmidt, Tamer Rabie, Ian Roberts, and Sandy Cairncross (2007) "Interventions to improve water quality for preventing diarrhoea: Systematic review and meta-analysis," *BMJ*, 334 (7597), 782.

Codex Alimentarius Commission (2022) "General Principles of Food Hygiene (CXC 1-1969),"Technical report, FAO/WHO, Rome, `https://openknowledge.fao.org/items/ce338d15-a83b-4e09-992b-9d553fa6589c`, Adopted 1969; amended 1999; revised 1997, 2003, 2020, 2022; editorial corrections 2011. Accessed 2025-10-02.

Cohen, Jessica and Pascaline Dupas (2010) "Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment," *The Quarterly Journal of Economics*, 125 (1), 1–45.

Cook, Annie, Kate Ambler, Vivian Hoffman, Lilian Otoigo, Alice Kiarie, and Julia Wagner (2025) "Capacity, regulatory oversight, and worker incentives in slaughter hygiene: Experimental evidence from Western Kenya," Working Paper.

Cortese, Rayza Dal Molin, Marcela Boro Veiros, Charles Feldman, and Suzi Barletto Cavalli (2016) "Food safety and hygiene practices of vendors during the chain of street food production in Florianopolis, Brazil: A cross-sectional study," *Food Control*, 62, 178–186.

Daniele, Gianmarco, Sulagna Mookerjee, and Denni Tommasi (2021) "Informational shocks and street-food safety: A field study in urban India," *Review of Economics and Statistics*, 103 (3), 563–579.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2008) "Returns to capital in microenterprises: Evidence from a field experiment," *The Quarterly Journal of Economics*, 123 (4), 1329–1372.

Deutschmann, Joshua W., Tanguy Bernard, and Ouambi Yameogo (2023) "Contracting and quality upgrading: Evidence from an experiment in Senegal," This version: January 19, 2023. Latest version available online.

Devoto, Florencia, Esther Duflo, Pascaline Dupas, William Parienté, and Vincent Pons (2012) "Happiness on tap: Piped water adoption in urban Morocco," *American Economic Journal: Economic Policy*, 4 (4), 68–99.

Duflo, Esther, Pascaline Dupas, Thomas Ginn, Grace Makana Barasa, Moses Baraza, Victor Pouliquen, and Vandana Sharma (2019) "HIV prevention among youth: A randomized controlled trial of voluntary counseling and testing for HIV and male condom distribution in rural Kenya," *PloS one*, 14 (7), e0219535.

Dupas, Pascaline (2009) "What matters (and what does not) in households' decision to invest in malaria prevention?" *American Economic Review Papers and Proceedings*, 99 (2), 224–30.

Dupas, Pascaline, Vivian Hoffmann, Michael Kremer, and Alix Peterson Zwane (2016) "Targeting health subsidies through a nonprice mechanism: A randomized controlled trial in Kenya," *Science*, 353 (6302), 889–895.

FAO (2007) *Promises and Challenges of the Informal Food Sector in Developing Countries*, New York: Food and Agriculture Organization.

FAO (2017) *Food Safety Risk Management: Evidence-Informed Policies and Decisions, Considering Multiple Factors* in , Food Safety and Quality Series (4), Rome, Italy: FAO.

FAO (2023) *Introduction and Control of Food Hazards – Section 1. FAO Good Hygiene Practices (GHP) and Hazard Analysis and Critical Control Point (HACCP) Toolbox for Food Safety*, Rome: Food and Agriculture Organization.

——— (2024) "Food Quality,"Technical report, https://www.fao.org/faoterm/viewentry/en/?entryId=170764, Definition of food quality: "The attributes of a food that influence its value and that make it acceptable or desirable for the consumer." Accessed 2025-10-02.

Fewtrell, Lorna, Rachel B Kaufmann, David Kay, Wayne Enanoria, Laurence Haller, and John M Colford Jr (2005) "Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: A systematic review and meta-analysis," *The Lancet: Infectious Diseases*, 5 (1), 42–52.

Fischer, Greg, Dean Karlan, Margaret McConnell, and Pia Raffler (2019) "Short-term subsidies and seller type: A health products experiment in Uganda," *Journal of Development Economics*, 137, 110–124.

Gargiulo, A. H., S. G. Duarte, G. Z. Campos, M. Landgraf, B. D. G. M. Franco, and U. M. Pinto (2022) "Food safety issues related to eating in and eating out," *Microorganisms*, 10 (11), 2118.

Hanna, Rema, Esther Duflo, and Michael Greenstone (2016) "Up in smoke: The influence of household behavior on the long-run impact of improved cooking stoves," *American Economic Journal: Economic Policy*, 8 (1), 80–114.

Hasanain, Syed Ali, Muhammad Yasir Khan, and Arman Rezaee (2023) "No bulls: Experimental evidence on the impact of veterinarian ratings in Pakistan," *Journal of Development Economics*, 161, 102999.

Haushofer, Johannes, Michael Kremer, Ricardo Maertens, and Brandon Joel Tan (2021) "Water treatment and child mortality: Evidence from Kenya," Working Paper, National Bureau of Economic Research.

Ho, Lisa Y, Emily Breza, Marcella Alsan et al. (2023) "The impact of large-scale social media advertising campaigns on COVID-19 vaccination: Evidence from two randomized controlled trials," *AEA Papers and Proceedings*, 113, 653–58.

Hoffmann, Vivian and Kelly Jones (2021) "Improving food safety on the farm: Experimental evidence from Kenya on incentives and subsidies for technology adoption," *World Development*, 143, 105406.

Hoffmann, Vivian, Sarah Kariuki, Janneke Pieters, and Mark Treurniet (2023) "Upside risk, consumption value, and market returns to food safety," *American Journal of Agricultural Economics*, 105 (3), 914–939.

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, 10.1016/j.worlddev.2019.104611.

Hsu, Eric and Anne Wambugu (2024) “Can informed buyers improve goods quality? Experimental evidence from crop seeds,” Working Paper.

ILO (2018) *Women and Men in the Informal Economy: A Statistical Picture. Third Edition*, Geneve, Switzerland: International Labor Organization.

Imai, Kosuke, Luke Keele, and Dustin Tingley (2010) “A general approach to causal mediation analysis.,” *Psychological methods*, 15 (4), 309.

Imai, Kosuke, Gary King, and Clayton Nall (2009) “The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation,” *Statistical Science*, 24 (1), 29–53.

Imbens, Guido W (2011) “Experimental design for unit and cluster randomid trials,”Technical report, International Initiative for Impact Evaluation Paper.

Jayachandran, Seema (2021) “Microentrepreneurship in developing countries,” in Zimmermann, Klaus F. ed. *Handbook of Labor, Human Resources and Population Economics*: Springer Nature, 10.1007/978-3-319-57365-6_174-1.

Kremer, Michael and Edward Miguel (2007) “The illusion of sustainability,” *The Quarterly Journal of Economics*, 122 (3), 1007–1065.

Kremer, Michael, Edward Miguel, Sendhil Mullainathan, Clair Null, and Alix Peterson Zwane (2011) “Social engineering: Evidence from a suite of take-up experiments in Kenya,” Working Paper.

Lane, Gregory, David Schonholzer, and Erin Kelley (2023) “Information and strategy in lemon markets: Improving safety in informal transit,” Working Paper.

Luoto, Jill, David Levine, Jeff Albert, and Stephen Luby (2014) “Nudging to use: Achieving safe water behaviors in Kenya and Bangladesh,” *Journal of Development Economics*, 110, 13–21.

Maestas, Nicole, Kathleen J. Mullen, David Powell, Till von Wachter, and Jeffrey B. Wenger (2023) “The value of working conditions in the United States and the implications for the structure of wages,” *American Economic Review*, 113 (7), 2007–47.

Mas, Alexandre and Amanda Pallais (2017) “Valuing alternative work arrangements,” *American Economic Review*, 107 (12), 3722–3759.

McFadden, Daniel (1973) “Conditional logit analysis of qualitative choice behavior,” *Frontiers in Econometrics*, 4, 105–142.

McFadden, Daniel and Kenneth Train (2000) “Mixed MNL models for discrete response,” *Journal of Applied Econometrics*, 15 (5), 447–470.

Michelson, Hope, Anna Fairbairn, Brenna Ellison, Annemie Maertens, and Victor Manyong (2021) “Misperceived quality: Fertilizer in Tanzania,” *Journal of Development Economics*, 148, 102579.

Miguel, Edward and Michael Kremer (2004) “Worms: Identifying impacts on education and health in the presence of treatment externalities,” *Econometrica*, 72 (1), 159–217.

Muinde, OK and E Kuria (2005) “Hygienic and sanitary practices of vendors of street foods in Nairobi, Kenya,” *African Journal of Food, Agriculture, Nutrition and Development*, 5 (1).

Muyanja, Charles, Leontina Nayiga, Namugumya Brenda, and George Nasinyama (2011) "Practices, knowledge and risk factors of street food vendors in Uganda," *Food Control*, 22 (10), 1551–1558.

Nizame, Fosiul A, Mahbub U Alam, Abdullah A Masud, Abul K Shoab, Aftab Opel, Md Khairul Islam, Stephen P Luby, and Leanne Unicomb (2019) "Hygiene in restaurants and among street food vendors in Bangladesh," *The American Journal of Tropical Medicine and Hygiene*, 101 (3), 566.

Nyhan, Brendan, Jason Reifler, Sean Richey, and Gary L Freed (2014) "Effective messages in vaccine promotion: A randomized trial," *Pediatrics*, 133 (4), e835–e842.

Park, Sangyoon, Zhaoneng Yuan, and Hongsong Zhang (2023) "Technology training, buyer-supplier relationship, and quality upgrading in an agricultural supply chain," *The Review of Economics and Statistics*, 1–46.

Revelt, David and Kenneth Train (2000) "Mixed logit with repeated choices: Households' choices of appliance Efficiency Level," *The Review of Economics and Statistics*, 82 (4), 647–657.

Rhee, Michelle, Mahamadou Sissoko, Sharon Perry, Willi McFarland, Julie Parsonnet, and Ogobara Doumbo (2005) "Use of insecticide-treated nets (ITNs) following a malaria education intervention in Piron, Mali: A control trial with systematic allocation of households," *Malaria Journal*, 4 (1), 1–9.

Ritter, Michael, Eveline Camille, Christophe Velcine, Rose-Kerline Guillaume, and Daniele Lantagne (2017) "Optimizing household chlorination marketing strategies: A randomized controlled trial on the effect of price and promotion on adoption in Haiti," *The American Journal of Tropical Medicine and Hygiene*, 97 (1), 271.

Samapundo, Simbarashe, Ruth Climat, Ramize Xhaferi, and Frank Devlieghere (2015) "Food safety knowledge, attitudes and practices of street food vendors and consumers in Port-au-Prince, Haiti," *Food Control*, 50, 457–466.

Samapundo, Simbarashe, TN Cam Thanh, Ramize Xhaferi, and Frank Devlieghere (2016) "Food safety knowledge, attitudes and practices of street food vendors and consumers in Ho Chi Minh city, Vietnam," *Food Control*, 70, 79–89.

Satpathy, Suchismita and Rohit Jha (2022) "Intermittent water supply in Indian cities: Considering the intermittency beyond demand and supply," *AQUA—Water Infrastructure, Ecosystems and Society*, 71 (12), 1395–1407.

Shukla, Pallavi, Hemant K Pullabhotla, and Kathy Baylis (2022) "Trouble with zero: The limits of subsidizing technology adoption," *Journal of Development Economics*, 158, 102920.

Singh, R. B., M. S. Haque, and A. Grover (2015) "Drinking water, sanitation and health in Kolkata Metropolitan City: Contribution towards urban sustainability," *Geography, Environment, Sustainability*, 8 (4), 64–81.

The New Indian Express (2023) "97 per cent households yet to get safe drinking water," *The New Indian Express*, https://www.newindianexpress.com/nation/2023/sep/25/97-per-cent-households-yet-to-get-safe-drinking-water-2618581.html.

The Telegraph India (2024) "Survey finds 60 per cent Indians use water filters, quality concerns rise amid heatwave," *The Telegraph India*, https://www.telegraphindia.com/india/survey-finds-60-per-cent-indians-use-water-filters-quality-concerns-rise-amid-heatwave/cid/2024450.

The Times of India (2019) "One-fifth of urban households in India use electric water purifiers, including ROs: NSO Survey," *The Times of India*, https://timesofindia.indiatimes.com/india/one-fifth-of-urban-households-in-india-use-electric-water-purifiers-including-ros-nso-survey/articleshow/72269209.cms.

Verhoogen, Eric (2023) "Firm-level upgrading in developing countries," *Journal of Economic Literature*, 61 (4), 1410–64, 10.1257/jel.20221633.

Vollaard, AM, S Ali, HAGH Van Asten, I Suhariah Ismid, S Widjaja, LG Visser, Ch Surjadi, and JT Van Dissel (2004) "Risk factors for transmission of foodborne illness in restaurants and street vendors in Jakarta, Indonesia," *Epidemiology & Infection*, 132 (5), 863–872.

WHO (2012) *How to Conduct a Discrete Choice Experiment for Health Workforce Recruitment and Retention in Remote and Rural Areas: A User Guide with Case Studies*, Geneva: World Health Organization.

——— (2022) *Food Safety*, Geneva: World Health Organization.

Wiswall, Matthew and Basit Zafar (2018) "Preference for the workplace, investment in human capital, and gender," *The Quarterly Journal of Economics*, 133 (1), 457–507.

Wongtada, Nittaya (2014) "Street vending phenomena: A literature review and research agenda," *Thunderbird International Business Review*, 56 (1), 55–75.

Wooldridge, Jeffrey M. (2023) "Simple approaches to nonlinear difference-in-differences with panel data," *Journal of Econometrics*.

World Bank (2019) *The Safe Food Imperative: Accelerating Progress in Low-and Middle-Income Countries*: World Bank Publications.

Online Appendix

- **Appendix A: Context – Details.** Provides background on the street food sector in Kolkata. Includes maps of study areas, photos of vendor kiosks, and tables on infrastructure, sanitation, licensing, electricity, and vendor finances.

- **Appendix B: Consumer Preferences for Safe Food – Details.** Presents full results from the discrete choice experiment. Includes effects of safety labels and cleanliness on consumer choices, and heterogeneity by income, gender, and cleanliness sensitivity.

- **Appendix C: Field Experiment with Vendors – Details.** Documents further the intervention design and logistics. Includes survey questions on food safety inputs and practices, and additional tables on compliance and infrastructure by treatment arm.

- **Appendix D: Minor Deviations from Pre-Analysis Plan.** Documents small deviations from the pre-analysis plan, including index adjustments and sample refinements made for feasibility or clarity.

- **Appendix E: Minor Errors in Data Collection.** Details minor coding errors in the data, in the spirit of transparency and reproducibility.

- **Appendix F: Estimation Strategy and Results.** Reports additional estimation results for large equipment, small supplies, and food safety practices. Includes index-level regressions, cumulative distribution plots, vendor-reported perceptions, chlorine use over time, and vendor-reported changes.

- **Appendix G: Robustness Checks and Treatment Effects Heterogeneity.** Analyzes heterogeneity by neighborhood, cluster size, and kiosk size. Includes interaction regressions and distributional effects.

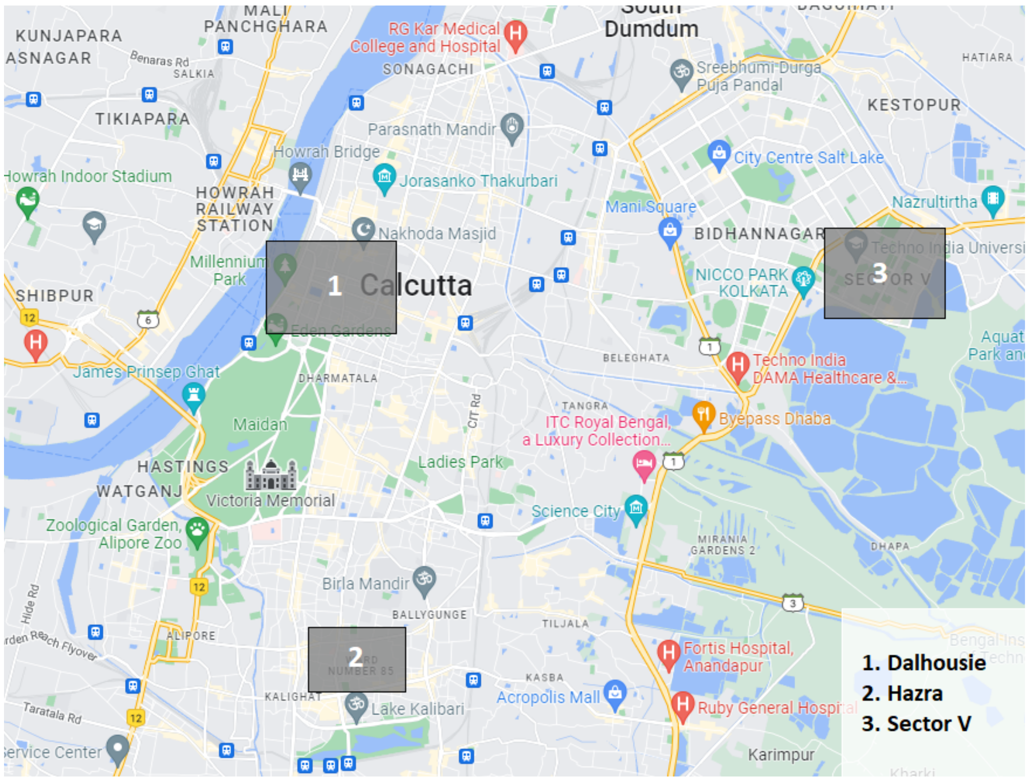
- **Appendix H: Robustness Checks for Business Outcomes.** Presents additional robustness checks for business outcomes. Confirms that core results are stable across alternative specifications.

- **Appendix I: Further Results on Spillover Effects.** Presents additional results on spillover effects for quality indices and business outcomes.

- **Appendix J: Conceptual Framework – Details.** Provides further details and discussion about the conceptual framework.

A Context: Details

Figure A1: Areas of Kolkata Included in the Sample



Notes: This figure shows a map of Kolkata with the 3 selected areas of our intervention. Our post-baseline sample consists of 274 vendors spread out across these 3 areas. Each area is represented by both treated and non treated vendors. There are 97 vendors in the control group (40 clusters), 92 in the first treatment (35 clusters) and 85 in the second treatment (33 clusters).

Figure A2: Example of a Street Vendor

Notes: This picture provides a graphical illustration of the typical vendor and kiosk in our sample.

Figure A3: Examples of Conditions of Kiosk Operations

Notes: Photos collected from vendors included in our sample during the pre-intervention period.

Table A1: Initial Context Assessment: Facilities (By Area)

	(1) Dalhousie	(2) Hazra	(3) Sector V
Kiosk has handwashing facility	0.95	0.96	0.96
Kiosk has garbage bin	0.86	0.88	0.93
Kiosk has drinking water facility	1.00	0.99	1.00
Handwashing facility has soap	0.36	0.59	0.47
Vendor uses an apron	0.03	0.10	0.19
Vendor wears gloves	0.00	0.00	0.00
Vendor wears hair cover	0.01	0.02	0.01
Treats water in primary storage	0.01	0.02	0.13

Notes: Data is from baseline pre-treatment surveys (surveys 1 and 2) conducted during our initial assessment of the context in May 2022. The table shows average values for each variable. See Section 4.4 for details on the data collection.

Table A2: Price-Quality Relationship

	(1)	(2)	(3)	(4)	(5)	(6)
		Asset Index			Food-Safety Practices Index	
Price	0.007	0.000	-0.002	-0.001	-0.004	-0.007
	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)	(0.006)
Area FE:	No	Yes	Yes	No	Yes	Yes
Meal FE:	No	No	Yes	No	No	Yes
Employee FE:	No	No	Yes	No	No	Yes
Control mean:	0.43	0.43	0.43	0.58	0.58	0.58
Clusters:	108	108	108	108	108	108
Observations:	1785	1785	1785	1785	1785	1785

Notes: Data is from baseline pre-treatment surveys (surveys 1 and 2) conducted during our initial context assessment in May 2022. See Section 4.4 for details on the data collection. "Asset index" is the average of the facilities in Table 1 across surveys 1 and 2; "Food-Safety Practices Index" is the average of the practices listed in Table 3 across surveys 1 and 2. "Price" is the price of the item sold by the vendor. Regressions are the item level. Meal FE controls for whether the kiosk produces meals or snacks. Employee FE controls for the number of employees working in the kiosk. Standard errors at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table A3: Vendor Desired Upgrades and Finances

	Obs.	Mean	S.D.	Min	Max
Primary way to spend 10,000₹ windfall for business:					
Maintenance, repair, and renovation of kiosk	274	0.49	0.50	0	1
Buying or upgrading equipment	274	0.24	0.43	0	1
Raw materials and items for resale	274	0.14	0.34	0	1
Introducing new food items	274	0.09	0.29	0	1
Repayment of loans	274	0.01	0.10	0	1
To hire more employees and apprentices	274	0.00	0.06	0	1
Finance:					
Has bank account	274	0.93	0.25	0	1
Use bank account	256	0.57	0.50	0	1
Ever applied for bank loan	274	0.16	0.36	0	1
No savings	274	0.53	0.50	0	1
Monthly savings (if positive)	126	5843.76	5609.93	167	28000

Notes: Data is from the baseline survey conducted in April 2022. In the windfall section, vendors were limited to selecting only one response. Monthly savings are in rupees. See Section 4.4 for details on the data collection. "Use bank account" is conditional on having a bank account. "Monthly savings" is conditional on having any savings.

B Consumer Preferences for Safe Food: Details

To estimate Willingness-to-Pay (WTP) using the choice experiment data, we follow standard practice by starting with a random utility model (McFadden, 1973). Assume that the utility a consumer i, derives from a specific alternative, labeled as a, within choice scenario t, can be expressed as follows:

$$U_{iat} = \mathbf{x}'_{iat}\boldsymbol{\beta}_i + \mathbf{w}'_{ia}\boldsymbol{\alpha} + \mathbf{z}'_i\boldsymbol{\delta}_a + \varepsilon_{iat} \tag{A1}$$

where $\boldsymbol{\beta}_i$ represents a vector of individual-specific coefficients, $\mathbf{x}_{iat}$ stands for a vector of alternative-specific variables, $\boldsymbol{\alpha}$ denotes fixed coefficients pertaining to $\mathbf{w}_{ia}$, a vector of alternative-specific variables, $\boldsymbol{\delta}_a$ signifies fixed alternative-specific coefficients for $\mathbf{z}_i$, a vector of consumer-specific variables, and ε_{iat} is a random term following a Type I extreme value distribution. Allowing $\boldsymbol{\beta}_i$ to vary among consumers accounts for the fact that different consumers may exhibit distinct preferences.

Following McFadden and Train (2000), the mixed logit choice probability is given by:

$$P_{iat} = \int \frac{\exp(\mathbf{x}'_{iat}\boldsymbol{\beta}_i)}{\sum_{j=1}^{J}\exp(\mathbf{x}'_{iat}\boldsymbol{\beta}_j)} f(\boldsymbol{\beta}|\boldsymbol{\theta})d\boldsymbol{\beta}$$

The integral represents the integration over the distribution of $\boldsymbol{\beta}$, where $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ is the density function, and $\boldsymbol{\theta}$ represents the vector of parameters that describe the characteristics of the distribution.[64] These parameters are typically estimated via simulated maximum likelihood techniques (Revelt and Train, 2000).

In the context of consumer's WTP for an attribute, the vector of coefficients $\boldsymbol{\beta}_i$ plays a central role in quantifying how changes in attribute levels impact consumer choices. The formulation of the WTP can be broken down into four steps. First, simulate draws of individual-specific coefficients $\boldsymbol{\beta}_i$ from the distribution $\boldsymbol{\beta}_i^* \sim f(\boldsymbol{\beta}|\boldsymbol{\theta})$, which represents different sets of coefficients for each consumer in our dataset. Second, for each set of simulated coefficients $\boldsymbol{\beta}_i^*$, calculate the choice probabilities P_{iat}^* for each alternative a in each choice scenario t using the utility representation. Third, for each set of simulated coefficients $\boldsymbol{\beta}_i^*$, calculate the MWP for the specific attribute a using $MWP_{iat}^a = -(\frac{\partial P_{iat}^*}{\partial \beta^a})/(\frac{\partial P_{iat}^*}{\partial \beta^{\text{price}}})$. Finally, calculate the expected MWP for attribute a by averaging the MWP values across all sets of simulated coefficients:

$$E[WTP^a] = -\frac{E[\beta^a]}{\beta^{price}}$$

This formula quantifies how much consumers are willing to pay for a change in attribute a while considering the fixed price coefficient.

[64] Hence, $\boldsymbol{\theta}$ controls the shape and variability of the distribution from which the individual-specific coefficients $\boldsymbol{\theta}_i$ are drawn. It is important to note that $\boldsymbol{\beta}$ is treated as a random variable in the mixed logit model, and its values are drawn from the distribution described by $\boldsymbol{\theta}$.

Figure A4: Example Question from the Discrete Choice Experiment

2 vendors selling a CHICKEN THALI (combo dish).
Both meals look equally tasty, but you have not tried either of them before.
Which option would you choose?

	Option A	**Option B**
Distance	Vendor is in front of you	Vendor is a 5 minute walk from you
Vendor's personal hygiene	Appears not very clean and hygienic	Appears very clean and hygienic
Kiosk's hygienic conditions	Appears not very clean and hygienic	Appears very clean and hygienic
Price	Rs 50	Rs 80

Notes: An example from the consumer DCE. The full set of questions as well as the survey manual is available upon request.

Table A4: Mixed Logit Estimates and Consumer Willingness to Pay by Attribute

	(1) Veg item	(2) Non-veg item	(3) Full sample
Coefficients:			
Price	-0.10***	-0.07***	-0.08***
	(0.00)	(0.00)	(0.00)
Clean kiosk	7.94***	6.71***	7.10***
	(0.29)	(0.22)	(0.18)
Clean vendor	2.91***	2.14***	2.40***
	(0.14)	(0.10)	(0.09)
Far location	-0.29***	-0.14***	-0.20***
	(0.05)	(0.04)	(0.03)
Standard deviations:			
Clean kiosk	5.47***	4.94***	5.16***
	(0.24)	(0.20)	(0.16)
Clean vendor	2.95***	2.34***	2.64***
	(0.14)	(0.11)	(0.10)
Far location	-0.56***	-0.19**	-0.40***
	(0.06)	(0.08)	(0.05)
Observations	48,312	48,312	96,624
Pseudo R-squared	0.530	0.478	0.503
Willingness-to-pay in ₹			
Clean kiosk (₹)	79.7	96.1	90.9
Clean vendor (₹)	29.1	30.6	30.7
Location > 5 min walk (₹)	-2.9	-2.0	-2.5

Notes: Data from consumer survey. The top panels report mixed logit estimates of coefficients and standard deviations for each attribution. The bottom panel reports implied WTP for safer street food options in Indian Rupees. Robust standard errors, clustered at the cluster level, are in parentheses. *p $<$ 0.10, **p $<$ 0.05, ***p $<$ 0.01.

Table A5: Correlations Between WTP and Demographic Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
	Veg item			Non-veg item		
	WTP kiosk hyg.	WTP vendor hyg.	WTP location	WTP kiosk hyg.	WTP vendor hyg.	WTP location
Hygiene is most important [0,1]	30.77***	8.78***	0.13	44.18***	10.43***	0.09
	(3.51)	(1.25)	(0.17)	(5.46)	(1.52)	(0.07)
Known contaminants [0,6]	1.39	0.12	-0.01	3.18	-1.50	0.00
	(1.63)	(0.63)	(0.06)	(1.89)	(0.80)	(0.03)
=1 if attended school	-2.68	8.17***	-0.25	25.18**	6.76	-0.25
	(7.20)	(2.25)	(0.41)	(9.30)	(3.54)	(0.27)
=1 if high-skill employee	8.18**	0.31	-0.11	14.56***	1.49	-0.07
	(2.43)	(1.20)	(0.13)	(3.17)	(1.35)	(0.05)
=1 if self-employed	-2.31	-1.56	-0.11	1.59	1.79	-0.05
	(2.85)	(1.34)	(0.19)	(4.21)	(1.92)	(0.10)
Dalhousie	2.78	-3.11*	0.02	-0.11	-2.41	0.05
	(3.55)	(1.45)	(0.18)	(6.16)	(2.15)	(0.07)
Sector V	7.22*	-1.06	0.16	8.78	0.06	0.05
	(3.58)	(1.57)	(0.22)	(6.72)	(2.27)	(0.08)
Control mean:	80.4	29.1	-3.0	96.7	30.3	-2.1
Clusters:	106	106	106	107	107	107
Observations:	1342	1342	1342	1342	1342	1342
Adjusted R^2:	0.29	0.19	0.01	0.41	0.19	-0.01

Notes: Data from the consumer survey. The dependent variables are individual Willingness-to-Pay (WTP) estimates for kiosk hygiene, vendor hygiene, and distance to the kiosk, derived separately from mixed logit models for items 1 and 2. The table reports OLS estimates of coefficients for a selected set of demographic characteristics, controlling for interviewer fixed effects. Standard errors are clustered at the cluster level. *p $<$ 0.10, **p $<$ 0.05, ***p $<$ 0.01.

Table A6: Conditional Logit Estimates and Consumer Willingness to Pay by Attribute

	(1) Veg item	(2) Non-veg item	(3) Full sample
Price	-0.035***	-0.027***	-0.029***
	(0.001)	(0.001)	(0.001)
Clean kiosk	2.362***	2.134***	2.231***
	(0.030)	(0.040)	(0.020)
Clean vendor	1.086***	0.835***	0.946***
	(0.026)	(0.022)	(0.017)
Far location	-0.102***	-0.069***	-0.084***
	(0.023)	(0.021)	(0.016)
Observations	48,312	48,312	96,624
Pseudo R-squared	0.530	0.478	0.503
Willingness-to-pay in ₹			
Clean kiosk (₹)	67.8	79.6	76.9
Clean vendor (₹)	31.1	31.1	32.6
Location > 5 min walk (₹)	-2.9	-2.6	-2.9

Notes: Data from consumer survey. The top panels report conditional logit estimates of coefficients for each attribution. The bottom panel reports implied WTP for safer street food options in Indian Rupees. Robust standard errors, clustered at the cluster level, are in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A7: Consumer Willingness to Pay by Area

	Veg item			Non-veg item			Both items		
	(1) Dalhousie	(2) Hazra	(3) Sector V	(4) Dalhousie	(5) Hazra	(6) Sector V	(7) Dalhousie	(8) Hazra	(9) Sector V
Price	-0.04***	-0.04***	-0.03***	-0.03***	-0.03***	-0.03***	-0.03***	-0.03***	-0.03***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Clean kiosk	2.28***	2.22***	2.77***	2.13***	1.91***	2.36***	2.19***	2.03***	2.53***
	(0.04)	(0.07)	(0.08)	(0.04)	(0.06)	(0.06)	(0.03)	(0.04)	(0.04)
Clean vendor	0.99***	1.20***	1.32***	0.80***	0.91***	0.86***	0.89***	1.03***	1.04***
	(0.03)	(0.06)	(0.07)	(0.03)	(0.05)	(0.05)	(0.02)	(0.04)	(0.04)
Far location	-0.10***	-0.08	-0.12**	-0.05*	-0.11**	-0.06	-0.08***	-0.09***	-0.09***
	(0.03)	(0.05)	(0.05)	(0.03)	(0.04)	(0.04)	(0.02)	(0.03)	(0.03)
Obs.:	26964	8820	12528	25740	9900	12672	52704	18720	25200
Pseudo R^2:	0.51	0.49	0.61	0.48	0.43	0.53	0.50	0.45	0.57
WTP kiosk hyg. (₹):	63.0	53.4	110.5	79.3	66.3	93.2	75.7	64.0	93.7
WTP vendor hyg. (₹):	27.4	28.9	52.7	29.7	31.7	34.1	30.7	32.5	38.5
WTP distance (₹):	-2.9	-1.8	-4.8	-2.0	-3.8	-2.4	-2.6	-2.9	-3.5

Notes: Data from consumer survey. The top panels report conditional logit estimates of coefficients for each attribution by area. The bottom panel reports implied WTP for safer street food options in Indian Rupees. Robust standard errors, clustered at the cluster level, are in parentheses. *p $<$ 0.10, **p $<$ 0.05, ***p $<$ 0.01.

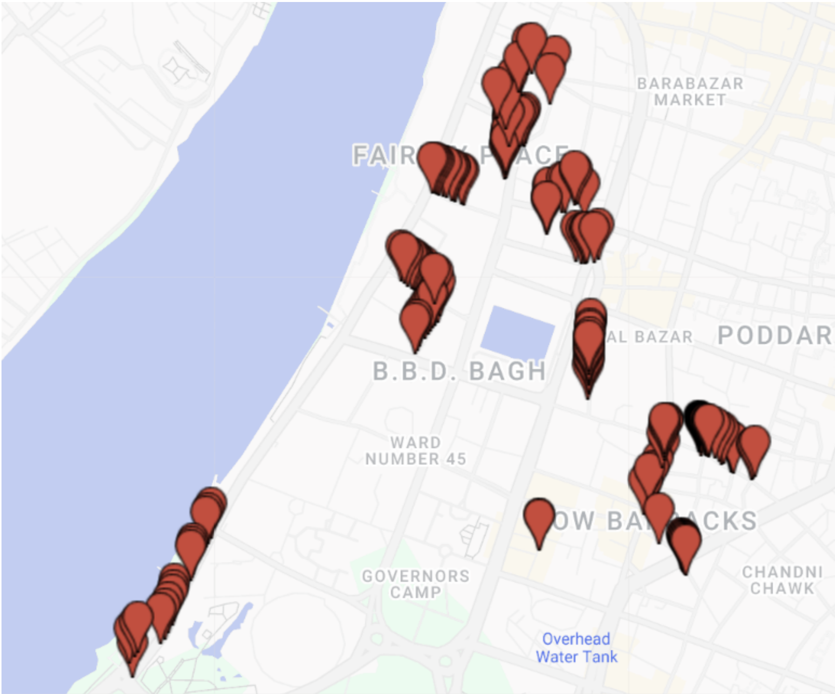
Table A8: Consumer Willingness to Pay by Demographic Group

	Veg Item			Non-veg Item			Both items		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Self	Low-skill	High-skill	Self	Low-skill	High-skill	Self	Low-skill	High-skill
Price	-0.04***	-0.04***	-0.02***	-0.03***	-0.03***	-0.02***	-0.03***	-0.04***	-0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Clean kiosk	2.07***	2.30***	2.78***	1.84***	1.96***	2.70***	1.94***	2.12***	2.73***
	(0.06)	(0.05)	(0.07)	(0.06)	(0.05)	(0.06)	(0.04)	(0.04)	(0.04)
Clean vendor	1.07***	1.00***	1.31***	0.88***	0.70***	1.04***	0.96***	0.84***	1.15***
	(0.06)	(0.04)	(0.06)	(0.05)	(0.04)	(0.05)	(0.04)	(0.03)	(0.04)
Far location	-0.15***	-0.05	-0.14***	-0.08	-0.06*	-0.08*	-0.11***	-0.06**	-0.11***
	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
Obs.:	8460	14580	14976	8568	12996	15948	17028	27576	30924
Pseudo R^2:	0.46	0.51	0.62	0.42	0.44	0.60	0.44	0.47	0.61
WTP kiosk hyg. (₹):	50.7	55.7	152.3	59.4	58.5	144.2	57.9	58.2	145.2
WTP vendor hyg. (₹):	26.2	24.2	71.4	28.3	20.8	55.6	28.8	23.0	61.4
WTP distance (₹):	-3.7	-1.2	-7.5	-2.5	-1.9	-4.5	-3.3	-1.6	-5.8

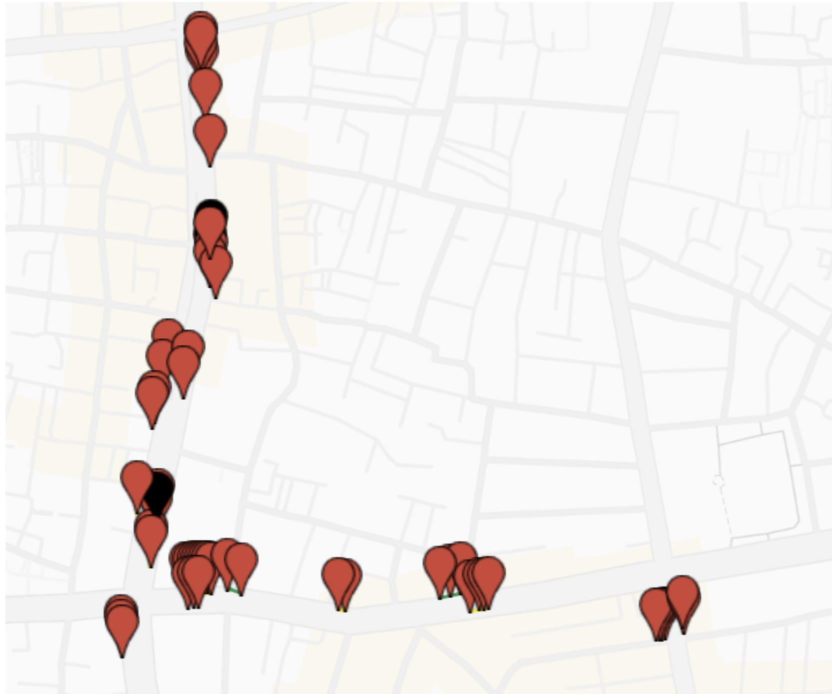
Notes: Data from consumer survey. The top panels report conditional logit estimates of coefficients for each attribution by selected demographic group. The bottom panel reports implied WTP for safer street food options in Indian Rupees. Robust standard errors, clustered at the cluster level, are in parentheses. *p $<$ 0.10, **p $<$ 0.05, ***p $<$ 0.01.

C Field Experiment with Vendors: Details

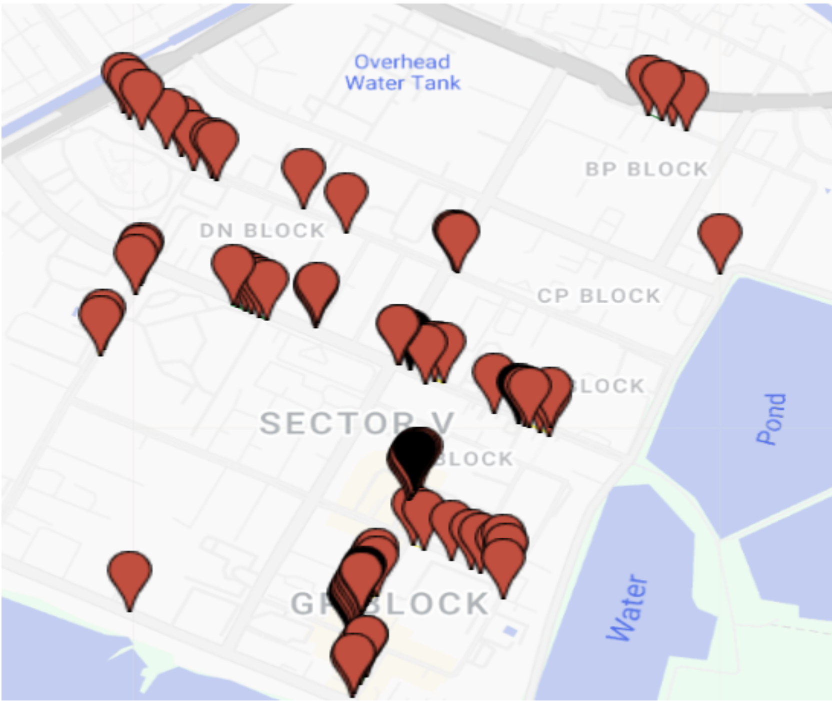
Figure A5: Distribution of Sampled Vendors



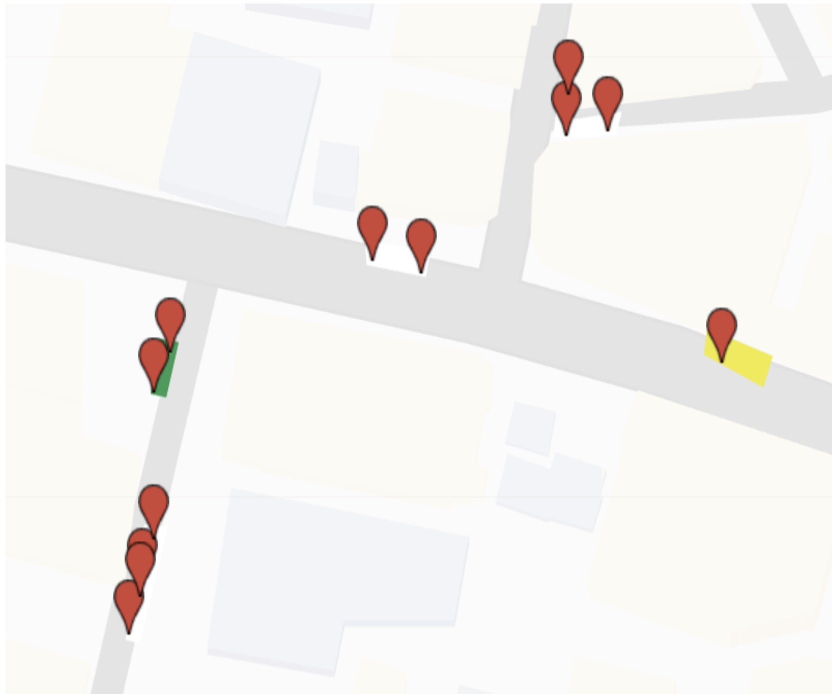
a) Distribution of vendors in Dalhousie



b) Distribution of vendors in Hazra



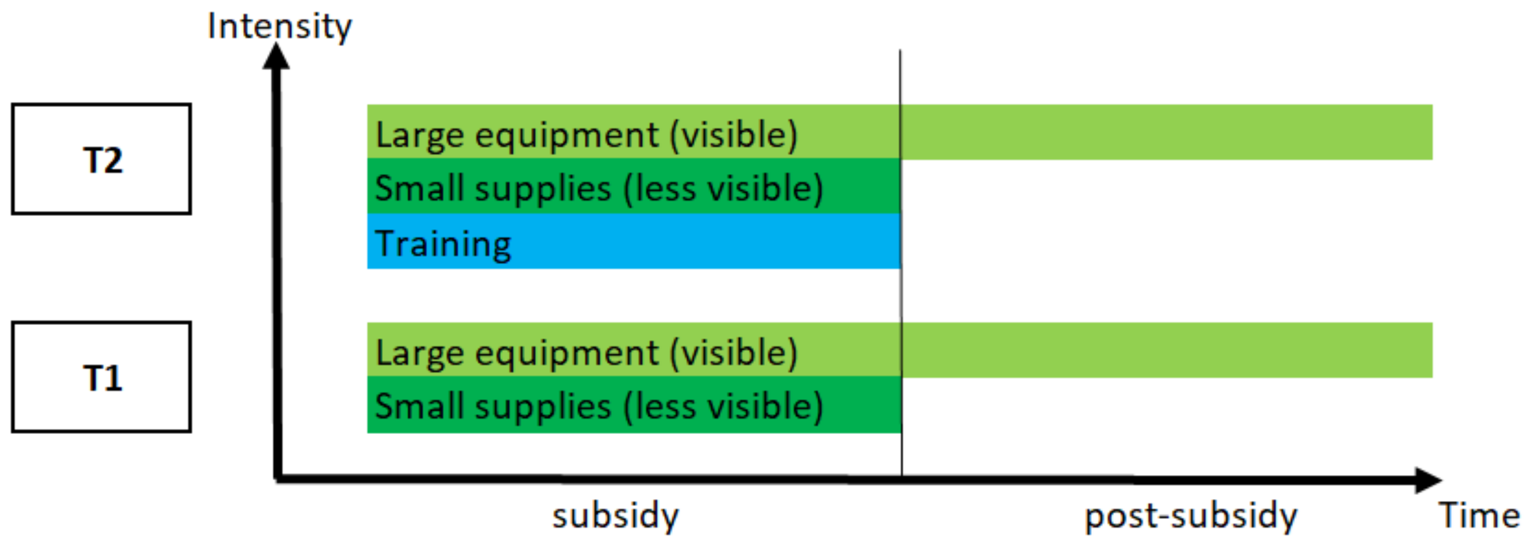
c) Distribution of vendors in Sector 5



d) Example street and clusters

Notes: Figures (a)–(c) show the distribution of sampled vendors in the Dalhousie area, Hazra, and Sector V, respectively. Our post-baseline sample consists of 274 vendors spread out across these 3 areas. Whereas, Figure (d) provides an example of vendors along a street and illustrates how we define natural urban clusters (each color block represents a cluster).

Figure A6: Summary of the Intervention

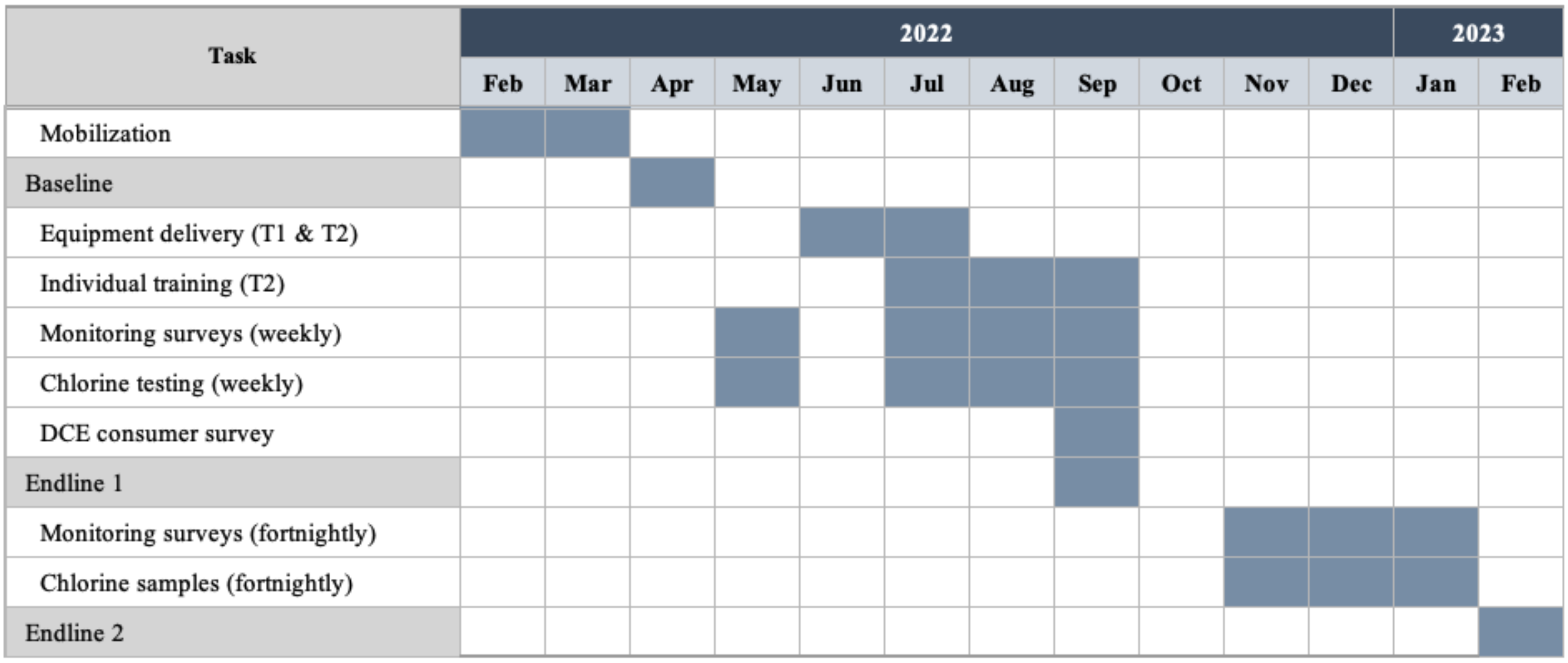


Notes: Timing and intensity of intervention components by treatment arm. All vendors in T1 and T2 received large visible equipment and weekly small supplies for 12 weeks starting in July 2023. Only T2 vendors received weekly training sessions during the same period. After the subsidy period, supplies were no longer provided for free, but vendors were informed where to purchase them locally.

Figure A7: Excerpt from Training Booklet for T2 Vendors

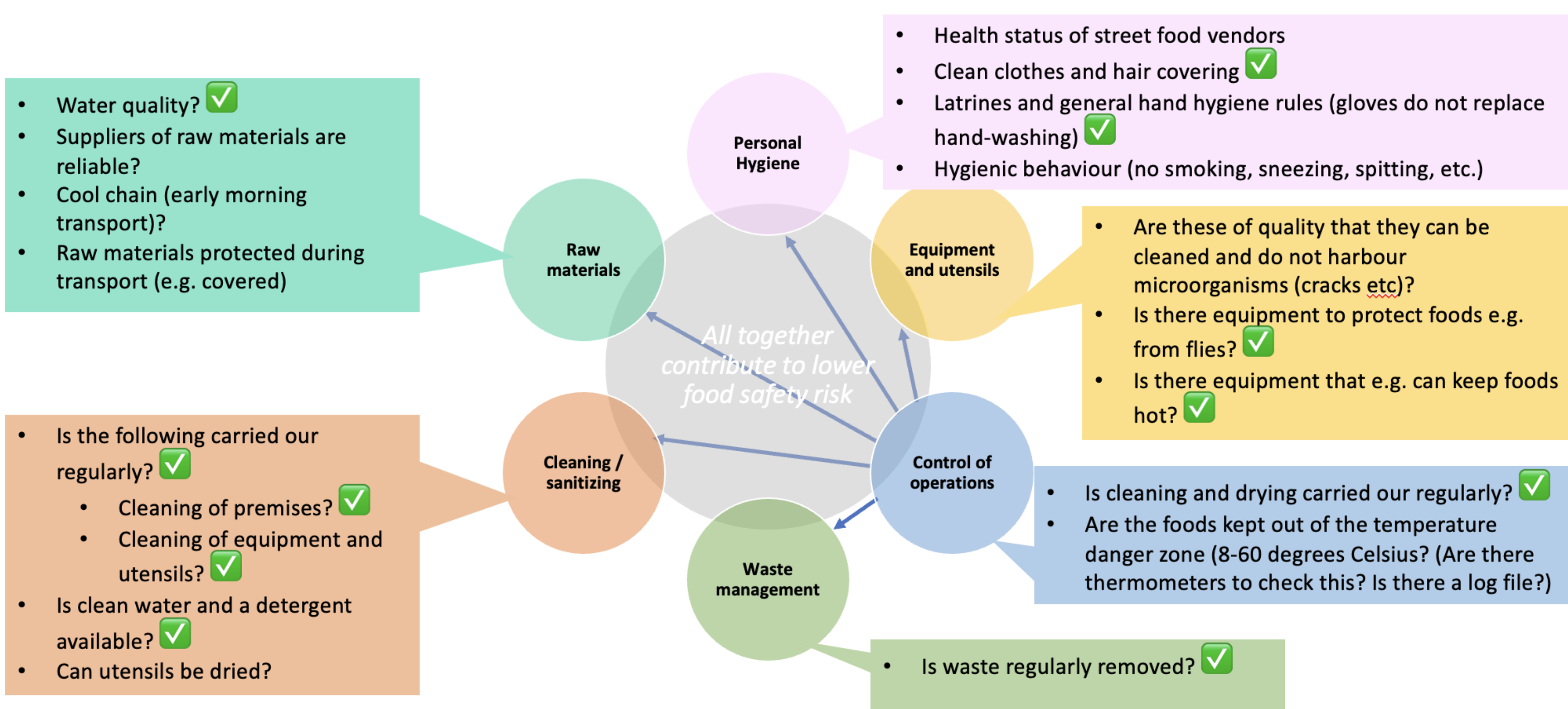
Notes: 12 golden rules for better, safer and hygienic street food (FSSAI).

Figure A8: Timeline of Data Collection



Notes: Project timeline for total study period, 2022-2023. Baseline data collection took place in April and May, 2022. Equipment was delivered over a three week period in late June/early July. Data collection continued throughout this period; our main results use data collected from mid-July when all large equipment had been delivered. Our first endline survey was administered at the end of September. Most vendors are away from their kiosks in October due to holidays; we did not collect data during this month. Post-treatment data collection began in November and continues fortnightly until the end of January. Our second endline was administered in mid February.

Figure A9: Framework for Constructing our Measure of Quality



Notes: Concept map of food safety domains and risks, developed in collaboration with food safety experts at FAO. Green checkmarks indicate the practices that were observable during field visits and included in our definition of quality. These practices, together with key sanitation-related inputs provided in the intervention, form the basis of the composite quality index used in the analysis.

Table A9: Survey Questions on Food Safety Inputs

Question	Possible Responses
Large Equipment	
1) Is there a visible facility for the vendor to wash hands?	**1. professional** 2. unprofessional 3. a mess
2) Does the stall have a garbage bin?	**1. Yes, it looks professional** 2. Yes, but it looks unprofessional 3. No
3) What does the primary water storage container look like?	**1. professional** 2. unprofessional 3. a mess
4) Is there a facility for drinking water for customers?	**1. Yes, it looks professional** 2. Yes, but it looks unprofessional 3. No
Small Supplies	
5) Is the vendor using an apron?	**1. yes, it looks clean and professional** 2. yes, but it looks dirty or unprofessional 3. no
6) Is the vendor wearing a hair cover?	**1. yes** 2. no
7) Is there soap available?	**1. Yes, it is full** **2. Yes, but almost empty** 3. No

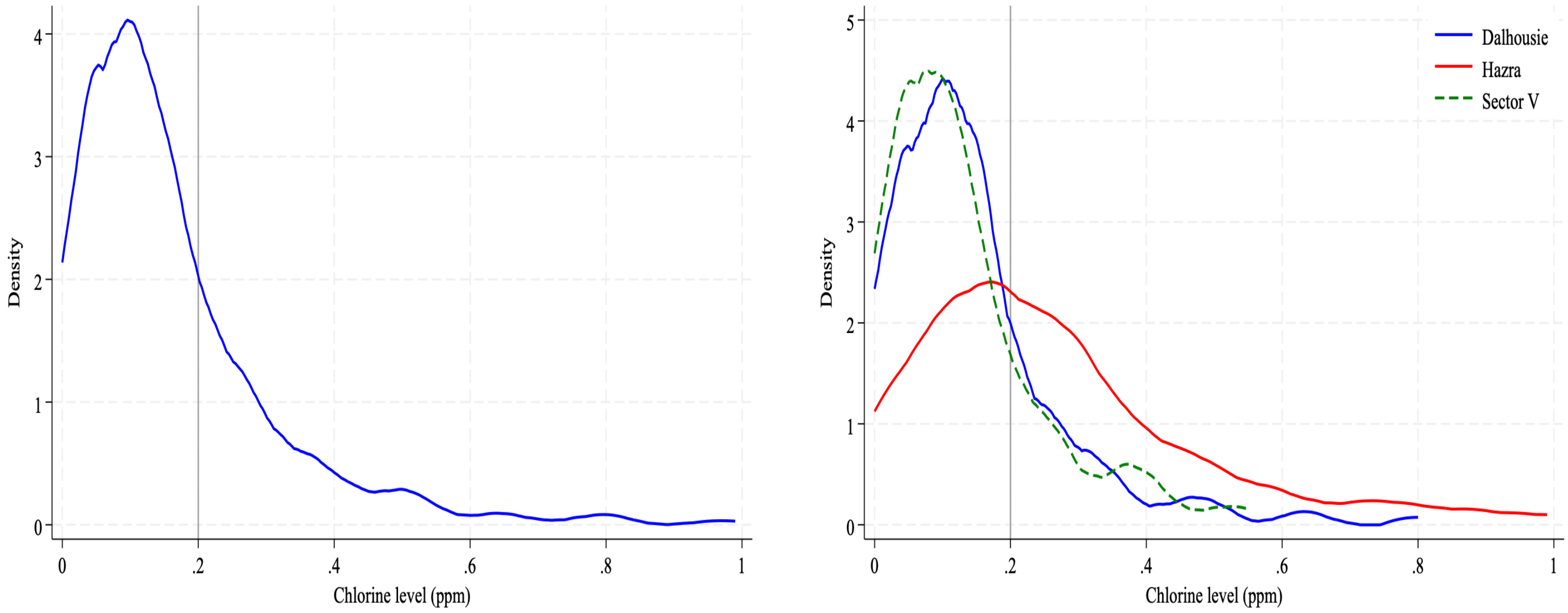
Notes: Questions from the Monitoring Survey on food-safety inputs. Responses in bold have been used to construct the indices and are coded as 0/1 indicators of "best behavior." These questions were developed in collaboration with food safety scientists at FAO, following the framework presented in Figure A9. They reflect practices consistent with the Codex Alimentarius and tailored to informal street food environments. For details on how these variables are used in the analysis, see Sections 4.1 and 5.1 in the main text.

Table A10: Survey Questions on Food Safety Practices

1) How does the water tank to clean the dishes look at the moment?	**1. has clean water** 2. has dirty water 3. is a mess 4. no tank/not visible
2) Does the vendor use soap to clean the dishes?	**1. Yes** 2. No 3. N/A - vendor not seen cleaning dishes but there is soap nearby 4. N/A - vendor not seen cleaning dishes and there is no soap nearby
3) Is there a towel/cloth for the vendor to wipe hands?	**1. yes, it looks clean and professional** 2. yes, but it looks dirty or unprofessional 3. no
4) How does the garbage bin look like at this moment?	**1. clean and empty** 2. with some garbage but not a mess 3. a mess
5) Is there visible garbage on the ground in or next to the kiosk?	1. Yes, quite a lot 2. Yes, a little **3. No, none**
6) Is there stagnant water on the ground in or next to the kiosk?	1. Yes, quite a lot 2. Yes, a little **3. No, none**
7) Is there raw or cooked food on the ground in or next to the kiosk?	1. Yes, quite a lot 2. Yes, a little **3. No, none**
8) Is the cooked food covered?	**1. yes, all food is fully covered (e.g. in a container with a lid)** **2. yes, covered behind a screen on the kiosk table** 3. yes, some food is fully covered but not all 4. not covered 5. N/A - not visible
9) Are the ingredients/raw food covered?	**1. yes, all of it is fully covered (e.g. in a container with a lid)** 2. yes, some of it is fully covered (e.g. in a container with a lid) 3. partially covered (e.g. with a cloth with holes) 4. not covered 5. N/A - not visible
10a) Are tongs/spatulas/other tools being used for cooking?	**1. yes** 2. no 3. yes, but also uses hands to touch food
10b) Is the food served with spoons, tongs or any other tools?	**1. yes** 2. no 3. yes, but also uses hands to touch food
11) Does the vendor wash hands before cooking/handling food?	**1. Yes, with soap** 2. Yes, without soap 3. No 4. Wears gloves 5. N/A - vendor not seen handling food
12) What does the counter where food is prepared look like?	**1. clean** 2. partially clean 3. dirty 4. very dirty
13) Is the food being served on disposable plates?	**1. Yes, all of them** 2. Yes, the majority of them 3. Yes, but only the minority of them 4. No, none of them

Notes: Questions from the Monitoring Survey on food-safety practices. Responses in bold have been used to construct the indices and enter in as a 0/1 variable. Question 10a) and 10b) have been combined. That is, variable 10 is equal to 1 if the vendor uses tongs or serving spoons for cooking or serving. Responses involving N/A are set to missing. These questions were developed in collaboration with food safety scientists at FAO, following the framework presented in Figure A9. They reflect practices consistent with the Codex Alimentarius and tailored to informal street food environments. For details on how these variables are used in the analysis, see Sections 4.1 and 5.1 in the main text.

Figure A10: Distribution of Chlorine Levels At Baseline



a) Full sample

b) By area

<u>Notes</u>: In Figure (a), we plot the distribution of chlorine levels across the entire vendor sample. In Figure (b), we plot the distribution of chlorine levels within each specific area. In both cases we use kernel density plots. Solid lines in both figures demarcate the recommended chlorine concentration boundaries for water (0.2ppm), as advised by the World Health Organization (WHO).

Table A11: Monitoring and Attrition

	Total		Control (T0)		Equipment (T1)		Equipment w/ training (T2)	
	Obs.	Attr.	Obs.	Attr.	Obs.	Attr.	Obs.	Attr.
Design								
# Areas	3		3		3		3	
# Clusters	108		40		35		33	
Monitoring and Attrition								
Pre-treatment (April to May 2022):								
Baseline	274	0.00	97	0.00	92	0.00	85	0.00
Period #1	252	0.08	87	0.10	86	0.07	79	0.07
Period #2	256	0.07	93	0.04	87	0.05	76	0.11
Treatment period (July to September 2022):								
Period #3	257	0.06	91	0.06	87	0.05	79	0.07
Period #4	263	0.04	92	0.05	88	0.04	83	0.02
Period #5	257	0.06	88	0.09	87	0.05	82	0.04
Period #6	263	0.04	92	0.05	90	0.02	81	0.05
Period #7	256	0.07	89	0.08	88	0.04	79	0.07
Period #8	256	0.07	88	0.09	87	0.05	81	0.05
Period #9	259	0.05	89	0.08	88	0.04	82	0.04
Period #10	258	0.06	91	0.06	86	0.07	81	0.05
Period #11	265	0.03	93	0.04	91	0.01	81	0.05
Endline 1	267	0.03	93	0.04	91	0.01	83	0.02
Post-treatment (November 2022 to February 2023):								
Period #12	260	0.05	91	0.06	89	0.03	80	0.06
Period #13	261	0.05	91	0.06	89	0.03	81	0.05
Period #14	259	0.05	91	0.06	88	0.04	80	0.06
Period #15	259	0.05	91	0.06	88	0.04	80	0.06
Period #16	261	0.05	91	0.06	89	0.03	81	0.05
Endline 2	255	0.07	89	0.08	86	0.07	80	0.06
Total	4938	0.05	1727	0.06	1677	0.04	1534	0.05

Notes: The table provides summary statistics of the monitoring surveys. Our post-baseline sample consists of 274 vendors, 97 in the control group (40 clusters), 92 in the first treatment (35 clusters) and 85 in the second treatment (33 clusters). The left (right) column of each panel shows the number of observations per monitoring period (attrition rates).

D Minor Deviations from Pre-Analysis Plan

We designed the questionnaires in close collaboration with food safety experts at FAO and refined them throughout the implementation of the field experiment. Each round of data collection was preceded by pre-testing and piloting in the field, conducted with support from our data collection partner. Still, given the complexity of conducting fieldwork in informal food settings in Kolkata, some unforeseen issues arose during implementation. This section documents the resulting deviations from the registered pre-analysis plan (PAP), with a focus on the outcome variables used in the paper.

In the PAP, we planned to group our questions to construct four separate indices: treatment infrastructure, kiosk facilities, food handling, and customer facilities. However, in our main analysis, we regroup the questions into three indices (plus an overall quality index, defined as the sum of all the items), based on refinements made during implementation. Specifically, we distinguish between "large equipment" and "small supplies" to separately capture the use of the different components of the treatment package, as we anticipated they might have distinct effects. We then merge the original food handling and kiosk environment indices into a single "food safety practices" index, which reflects both the effect of the training and any potential spillover from the provision of equipment or supplies. For example, vendors in the treatment groups might wash their hands more frequently with soap, given that soap was part of the treatment package. This structure helps streamline the analysis while providing a more comprehensive picture of behavioral change. Nevertheless, we report results for each component separately, and our main conclusions remain unchanged when we reconstruct and analyze the original indices defined in the PAP (results available upon request).

Table A12 provides a precise mapping between the indices proposed in the PAP and those used in the final analysis. The main differences are as follows. First, during the study period, it became clear that in many cases vendors and customers were using the same facilities (e.g., the same garbage bin or handwashing station). As a result, we combined or dropped questions related to customer-specific facilities. Second, we removed questions about disposable gloves, as the treatment was modified to provide new aprons instead. This decision, made in consultation with FAO, was based on concerns that gloves might reduce food safety by encouraging less frequent handwashing and contribute to waste. Third, two questions (2 and 11, which were duplicates) were removed due to a coding error in the questionnaire that caused them to be asked incorrectly in the field. We replaced these with Question 34, which addresses the same issue. We do not consider this a serious concern, as both vendors and customers used the same handwashing facility. Fourth, Question 31 was dropped because its response options did not allow us to clearly define a "best behavior" in this context. Lastly, Questions 16 and 20 were dropped due to lack of variation.

Regarding interpretation of results, the PAP also specified that we would examine vendor knowledge of food safety practices (Hypothesis 2a). This was ultimately incorporated into our analysis of the food safety training's effectiveness (see Section 5.3). Here, we deviate from the PAP by excluding two questions: *"How frequently is water used to wash dishes and utensils changed?"* and *"How frequently is the main garbage bin emptied?"* We excluded these because the response options did not allow us to clearly identify the appropriate benchmark behavior. We also excluded the question about what vendors think is best for handwashing, as all vendors answered "soap," leaving no variation.

Table A12: Explanation of Deviation from Pre-Analysis Plan Proposed Indices

Item	Survey Question	Index/Comments
Treatment infrastructure		
1	Is there a visible facility for the vendor to wash hands?	Large equipment
2	Is there soap visible nearby?	Removed as there was an issue with how this question was asked by enumerators
3	Is the vendor wearing gloves?	Removed as we decided not include gloves with small supplies
4	Is the vendor wearing a hair cover?	Small supplies
5	Does the stall have a garbage bin?	Large equipment
6	Is there a garbage bin for consumers to dispose of waste?	Removed because vendors and customers use the same bin
7	Is there a facility for drinking water for customers?	Large equipment
8	Is there a hand washing facility for customers?	Large equipment
9	Is there soap available?	Small supplies
Kiosk facilities		
10	Is there a visible facility for the vendor to wash hands?	Large equipment
11	Is there soap visible nearby?	Removed as there was an issue with how this question was asked by enumerators
12	How does the water tank to clean the dishes look?	Practices
13	Whether there is soap visible in or next to the water?	This question changed to "Does vendor use soap to clean dishes" and added to Practices
14	Is there a towel/cloths for the vendor to wipe hands?	Practices
15	Is the vendor using an apron?	Small supplies
16	Is the vendor wearing gloves?	Removed from the index as 99% of vendors do not use gloves
17	Is the vendor wearing a hair cover?	Small supplies
18	Does the stall have a garbage bin?	Large equipment
19	How does the garbage bin look like at this moment?	Practices
Food handling		
20	Are the ingredients, raw or half-cooked food separated from cooked food?	Removed from the index as almost all vendors did separate
21	Is the cooked food covered?	Practices
22	Are the ingredients/raw food covered?	Practices
23	Are tongs/spatulas/other tools being used for cooking?	Practices
24	Is the food served with spoons, tongs or any other tools?	Practices
25	Does the vendor wash hands before cooking/handling food?	Practices
26	What does the counter where food is prepared look like?	Practices
Customer facilities		
27	Is the food being served on disposable plates?	Practices
28	Is there a garbage bin for consumers to dispose of waste?	Removed because vendors and customers use the same bin
29	How does the garbage bin look like at this moment?	Removed because vendors and customers use the same bin
30	Is there a facility for drinking water for customers?	Large equipment
31	How is drinking water distributed?	Removed because it was unclear which of the possible responses was "best behavior"
32	Is there a hand washing facility for customers?	Removed because vendors and customers use the same facility
33	How does this facility look at the moment?	Removed because vendors and customers use the same facility
34	Is there soap available?	Small supplies

Notes: Responses to the survey questions are presented in Tables A9 and A10 in the Appendix. Questions 2 and 11 were asked incorrectly by enumerators due to a coding error in the questionnaire. For soap usage, we rely instead on responses to Question 9, which refers to the presence of soap near the customer hand-washing facility. We exclude the question on whether the vendor uses disposable gloves, as 99% of vendors do not. Similarly, we exclude the question on whether vendors separate raw from cooked food, since 99% of vendors report doing so. Finally, we exclude the question on how drinking water is distributed, as the available response options—(1) served in disposable glasses, (2) served in non-disposable glasses, or (3) served directly from a communal container—did not allow us to clearly define what constituted "best behavior" in this context.

E Minor Errors in Data Collection

While every effort was made to ensure accurate data collection, a few issues regarding secondary outcomes did arise during the study period. However, none of these affect the core variables used in the main analysis or compromise the validity of the results presented in the paper. We document them nonetheless in the spirit of transparency and reproducibility.

During Endline 1, a number of questions meant for treatment group vendors—concerning the equipment received and, for T2 vendors, the training—were inadvertently omitted due to an error in the questionnaire provided to the data collection team. As a result, the following questions were not asked for the primary water storage container, drinking water facility, hand-washing facility, and garbage bin:

- *"Over the next 3 months, will you continue to use the item that you received?"*
- *"What do you plan to do with the item that you received?"*
- *"What did you do with the item that you received?"*

The first of these questions was also not asked for aprons and hairnets. A similar issue occurred with the following question, which was not asked for soap and chlorine tablets:

- *"Over the next 3 months, will you be able to buy the item in the local market on your own?"*

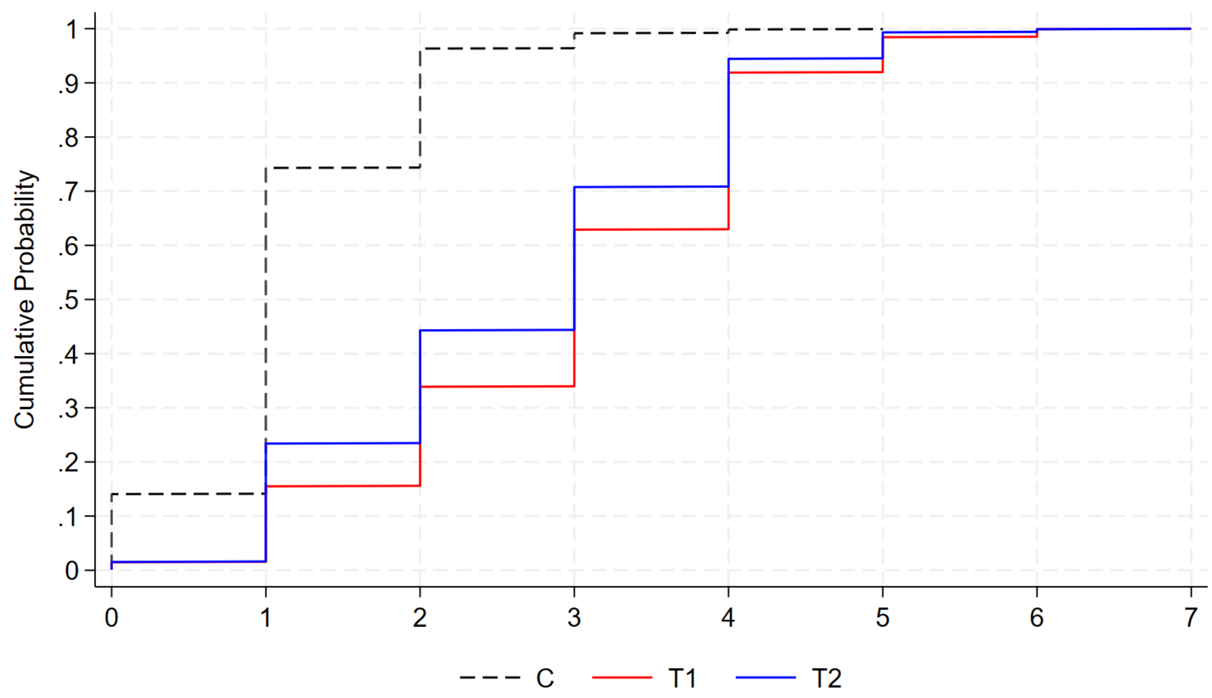
Additionally, for the training module, we were unable to ask: *"What changes did you make to your daily operations?"* To address these omissions, we included the relevant questions in the first post-treatment monitoring survey.

In the baseline survey, customer numbers were recorded using categories that severely underestimated actual customer counts, as it turned out that most vendors served more than the upper bound of the response options. Consequently, we exclude this variable from the baseline analysis and instead rely on customer counts from all subsequent surveys, which used a more appropriate scale.

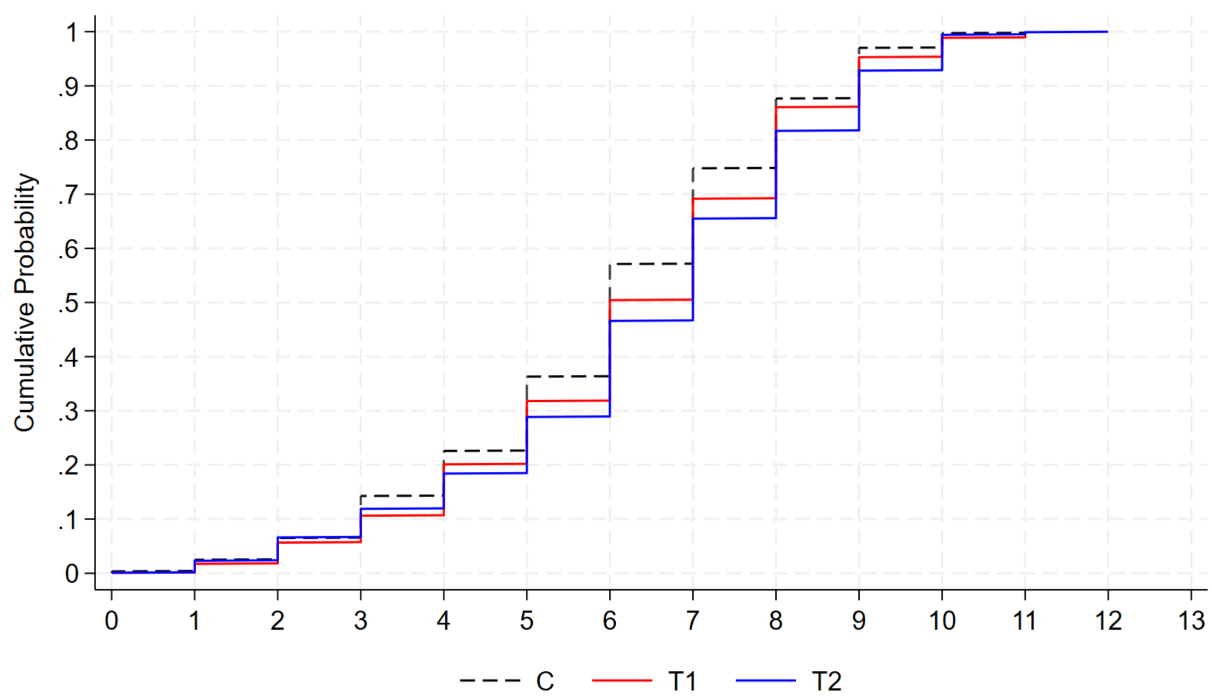
Finally, we initially collected price data every two weeks. However, we found that prices rarely changed between weeks. To streamline data collection, we discontinued the collection of vendor-reported price data after Monitoring Round 9, and shifted our focus to a detailed item-price census conducted at endline. As a result, our analysis of prices relies solely on baseline and endline data.

F Estimation Strategy and Results: Details

Figure A11: Treatment Effects on Large/Small Items and Food-Safety Practices



a) Large equipment and small supplies



b) Food-Safety Practices

<u>Notes</u>: Data from monitoring surveys (random audits) measured from period 5 (when the large equipment had been delivered to all vendors) up until the second endline survey in February 2023. Figure (a) plots the empirical cumulative distribution functions of the large equipment and small supplies count variable (0–7). Figure (b) plots the empirical cumulative distribution functions of the food-safety practices count variable (0–13).

Table A13: Treatment Effects on Large Equipment and Small Supplies By Index Component

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Large equipment				Small supplies		
	Handwash facility	Water container	Drinking water	Garbage bin	Clean apron	Hair cover	Handwash soap
Equipment (T1)	0.493***	0.213***	0.374***	0.703***	0.109***	0.060***	-0.051*
	(0.058)	(0.040)	(0.044)	(0.042)	(0.028)	(0.013)	(0.021)
	[0.001]	[0.001]	[0.001]	[0.001]	[0.002]	[0.001]	[0.046]
w/ training (T2)	0.410***	0.210***	0.257***	0.633***	0.066*	0.017*	-0.014
	(0.073)	(0.039)	(0.052)	(0.050)	(0.026)	(0.008)	(0.025)
	[0.001]	[0.001]	[0.001]	[0.001]	[0.036]	[0.057]	[0.439]
Equipment (T1) x post	0.109	-0.077	-0.249***	-0.114*	-0.087**	-0.040*	0.095**
	(0.074)	(0.041)	(0.047)	(0.046)	(0.030)	(0.016)	(0.033)
	[0.189]	[0.110]	[0.001]	[0.038]	[0.017]	[0.036]	[0.016]
Equipment (T2) x post	0.106	-0.080	-0.195***	-0.090*	-0.092**	-0.027***	0.021
	(0.067)	(0.044)	(0.051)	(0.036)	(0.028)	(0.008)	(0.032)
	[0.170]	[0.110]	[0.002]	[0.036]	[0.007]	[0.005]	[0.439]
Control mean:	0.08	0.79	0.02	0.06	0.05	0.00	0.25
Implied T1 effect (%):	644.7	26.9	1692.0	1155.1	208.3	6544.7	-20.3
Implied T2 effect (%):	536.6	26.4	1161.9	1041.4	125.1	1871.8	-5.4
Clusters:	108	108	108	108	108	108	108
Observations:	3073	3066	3073	3073	3073	3073	3041
Adjusted R^2:	0.45	0.31	0.28	0.48	0.18	0.08	0.57
p-value T1-T2:	0.30	0.88	0.04	0.16	0.21	0.01	0.10

Notes: Data from monitoring surveys (random audits) measured from period 5 (when the large equipment had been delivered to all vendors) up until the second endline survey in February 2023. Outcomes are binary variables taking value one if "Best behaviour" in the described category is observed, and zero otherwise. "post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. All OLS regressions include strata fixed effects and the pre-treatment average sanitary infrastructure observed at the kiosk. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors are always clustered at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A14: Treatment Effects on Practices By Index Component (Part 1)

	(1) Uses soap for dishes	(2) Dish water clean	(3) Garbage bin clean	(4) Garbage on ground	(5) Water on ground	(6) Food on ground
Equipment (T1)	0.019	0.058	0.209***	0.043	0.068	-0.029
	(0.010)	(0.033)	(0.046)	(0.043)	(0.049)	(0.035)
	[0.110]	[0.124]	[0.001]	[0.316]	[0.208]	[0.377]
w/ training (T2)	0.021**	0.051	0.180***	0.018	-0.009	-0.061
	(0.008)	(0.033)	(0.043)	(0.042)	(0.043)	(0.039)
	[0.029]	[0.172]	[0.001]	[0.472]	[0.568]	[0.170]
Equipment (T1) x post	-0.008	0.024	0.069	-0.051	-0.085	0.028
	(0.013)	(0.049)	(0.055)	(0.043)	(0.055)	(0.059)
	[0.439]	[0.458]	[0.249]	[0.268]	[0.172]	[0.458]
Equipment (T2) x post	-0.007	0.044	0.061	-0.038	-0.030	0.031
	(0.012)	(0.043)	(0.048)	(0.046)	(0.052)	(0.068)
	[0.439]	[0.316]	[0.246]	[0.377]	[0.439]	[0.467]
Control mean:	0.99	0.75	0.05	0.25	0.30	0.75
Implied T1 effect (%):	1.9	7.7	407.0	16.9	22.6	-3.8
Implied T2 effect (%):	2.1	6.8	351.1	7.3	-2.9	-8.1
Clusters:	108	108	108	108	108	108
Observations:	2044	2938	2706	3044	3040	3030
Adjusted R^2:	0.04	0.39	0.40	0.37	0.40	0.51
p-value T1-T2:	0.78	0.85	0.50	0.54	0.08	0.41

Notes: Data from monitoring surveys (random audits) measured from period 5 (when the large equipment had been delivered to all vendors) up until the second endline survey in February 2023. Outcomes are binary variables taking value one if "Best behaviour" in the described category is observed, and zero otherwise. "post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. All OLS regressions include strata fixed effects and the pre-treatment average sanitary infrastructure observed at the kiosk. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors are always clustered at the cluster level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A15: Treatment Effects on Practices By Index Component (Part 2)

	(1) Disposable plates	(2) Clean towel	(3) Counter clean	(4) Cooked food covered	(5) Raw food covered	(6) Use tongs or spoons	(7) Wash hands with soap
Equipment (T1)	0.044	0.024	0.014	0.095	0.078*	-0.018	-0.000
	(0.058)	(0.026)	(0.035)	(0.056)	(0.033)	(0.030)	(0.012)
	[0.408]	[0.358]	[0.487]	[0.145]	[0.046]	[0.439]	[0.646]
w/ training (T2)	0.094	0.015	0.013	0.140*	0.129**	-0.021	0.021
	(0.051)	(0.026)	(0.032)	(0.059)	(0.041)	(0.034)	(0.013)
	[0.110]	[0.439]	[0.487]	[0.046]	[0.008]	[0.439]	[0.150]
Equipment (T1) x post	-0.033	0.007	0.032	-0.103	-0.035	0.013	-0.003
	(0.047)	(0.035)	(0.043)	(0.074)	(0.039)	(0.062)	(0.014)
	[0.438]	[0.568]	[0.422]	[0.207]	[0.362]	[0.568]	[0.568]
Equipment (T2) x post	-0.027	0.035	0.038	-0.147*	-0.090*	-0.020	-0.034*
	(0.050)	(0.040)	(0.046)	(0.072)	(0.042)	(0.059)	(0.016)
	[0.439]	[0.362]	[0.377]	[0.079]	[0.069]	[0.523]	[0.065]
Control mean:	0.53	0.60	0.77	0.57	0.16	0.81	0.02
Implied T1 effect (%):	8.3	3.9	1.8	16.7	49.2	-2.2	-2.5
Implied T2 effect (%):	17.5	2.6	1.7	24.6	81.7	-2.5	126.6
Clusters:	108	108	108	108	108	108	108
Observations:	3073	3073	3073	3073	2642	3073	3020
Adjusted R^2:	0.42	0.63	0.36	0.28	0.46	0.41	0.05
p-value T1-T2:	0.33	0.75	0.97	0.54	0.22	0.93	0.14

Note: Data from monitoring surveys (random audits) measured from period 5 (when the large equipment had been delivered to all vendors) up until the second endline survey in February 2023. Outcomes are binary variables taking value one if "Best behaviour" in the described category is observed, and zero otherwise. "post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. All OLS regressions include strata fixed effects and the pre-treatment average sanitary infrastructure observed at the kiosk. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors are always clustered at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A16: Vendor Perceptions of Equipment Value

	(1)	(2)	(3)	(4)	(5)	(6)
	Use regularly		Adds value to business		Customer values	
	Total Mean/(SE)	Pairwise t-test (T2-T1)	Total Mean/(SE)	Pairwise t-test (T2-T1)	Total Mean/(SE)	Pairwise t-test (T2-T1)
Large equipment:						
Handwash facility	0.872 (0.031)	1.511	0.889 (0.030)	1.703*	0.873 (0.030)	1.875*
Water container	0.970 (0.009)	-0.781	0.982 (0.008)	-0.561	0.967 (0.010)	-0.265
Drinking water	0.883 (0.022)	-0.028	0.902 (0.022)	-0.271	0.883 (0.022)	-0.028
Garbage bin	0.946 (0.015)	0.970	0.952 (0.015)	0.725	0.939 (0.016)	0.944
Small equipment:						
Apron	0.381 (0.033)	-0.017	0.424 (0.033)	1.009	0.405 (0.036)	-0.016
Hair cover	0.263 (0.029)	-0.683	0.270 (0.029)	-0.856	0.260 (0.028)	-0.532
Handwash soap	0.761 (0.030)	0.062	0.773 (0.030)	0.240	0.728 (0.037)	1.405
Chlorine tablets	0.413 (0.019)	-0.452	0.396 (0.018)	-0.489	0.346 (0.021)	-0.515

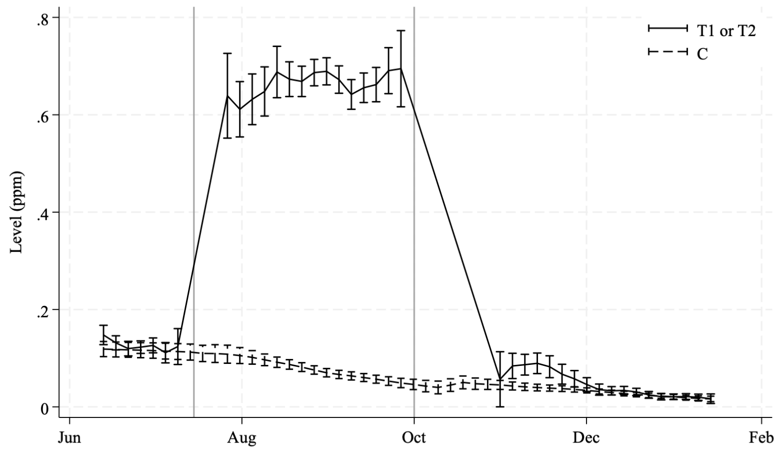
Notes: Data from Endline 1. Only treatment group vendors are included. Columns (1), (3) and (5) present the parameter estimates and the associated standard errors for the excluded group dummy, derived from a regression of the left variable on strata dummies and treatment dummies. Columns (2), (4) and (6) report the t-statistics from the difference tests. “Use regularly” refers to whether or not the vendor uses the equipment provided regularly; “Adds value to business” indicates whether or not the vendor thinks that the equipment adds value to the kiosk; “Customer values” refers to whether or not the vendor believes that customers value the equipment. Strata fixed and clustered standard errors used when comparing means. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table A17: Self-Reported Vendor Food Safety Knowledge and Behavior

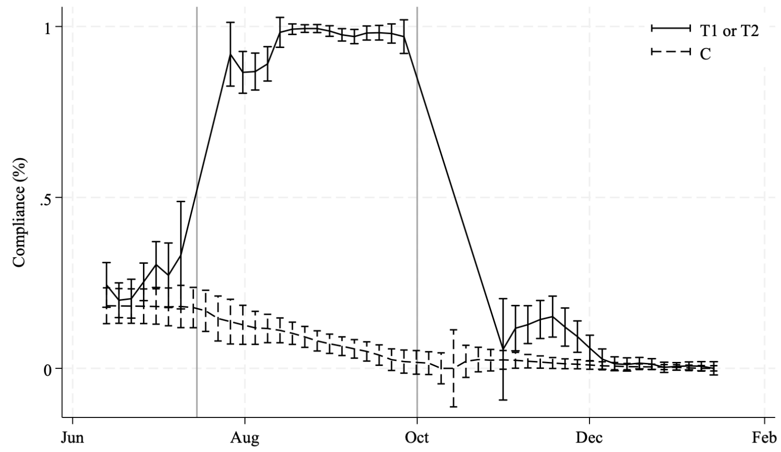
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			How easy it is for vendors to:				
	Reasons to wash hands	Treats water	Provide clean water	Empty garbage	Wash hands	Cover food	Change dish water
Equipment (T1)	0.007	0.113***	0.070	-0.011	0.007	-0.027	0.026
	(0.006)	(0.020)	(0.037)	(0.011)	(0.006)	(0.018)	(0.016)
	[0.223]	[0.001]	[0.146]	[0.223]	[0.223]	[0.203]	[0.150]
w/ training (T2)	0.019**	0.101***	0.041	-0.040**	-0.012	-0.011	-0.025
	(0.007)	(0.021)	(0.035)	(0.013)	(0.007)	(0.021)	(0.019)
	[0.023]	[0.001]	[0.223]	[0.008]	[0.150]	[0.334]	[0.223]
Control mean:	0.43	0.06	3.42	4.29	4.30	2.82	4.07
T1 effect (%):	0.8	12.0	7.2	-1.1	0.7	-2.7	2.7
T2 effect (%):	2.0	10.7	4.1	-4.0	-1.2	-1.1	-2.5
Clusters:	108	108	108	108	108	108	108
Observations:	2573	2573	514	514	514	514	514
p-value T1-T2:	0.09	0.57	0.35	0.04	0.01	0.44	0.00

Note: Data from monitoring and endline surveys. Column (1) is an index variable indicating how many correct reasons a vendor can identify when hand-washing is necessary. Column (2) is an indicator variable equal to 1 if the vendor reports treating the water in his primary water storage container. The remaining columns are ordinal variables from 1-5 where 1 indicates very difficult and 5 indicates very easy. We use OLS estimation for columns (1) and (2) and a Poisson regression model for estimation of columns (3)-(7). All regressions include strata fixed effects and the pre-treatment average sanitary equipment observed at the kiosk. To increase precision, we also include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Figure A12: Chlorine Usage over Time



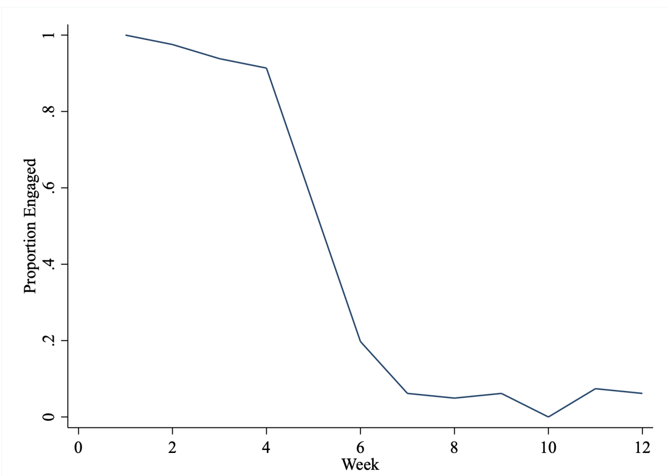
a) Chlorine level (ppm)



b) Chlorine compliance (%)

Notes: Data from monitoring surveys (random audits). Figure a) presents the distribution of chlorine presence (in ppm) in primary water storage containers. Figure b) compares the percentage of vendors with free chlorine levels in the water above 0.20 ppm, which is the minimum requirement, during the observational period.

Figure A13: Proportion of Vendors Considered Engaged in Training Sessions



Notes: Data from evaluations of sessions by trainers. Trainers evaluated the vendor's engagement following each training session, using a Likert scale to assess engagement. We consider vendors as being engaged if they are reported as being engaged or very engaged on the Likert scale. Only T2 vendors included; 891 observations collected over a 12 week period.

Table A18: Vendor Perceptions of Training and Issues with Implementation

	Obs.	Mean	S.D.	Min	Max
Perceptions of the training:					
Training was useful	74	0.26	0.44	0	1
Would recommend training to other vendors	74	0.26	0.44	0	1
Training changed how kiosk is run	74	0.27	0.45	0	1
Training adds value to the kiosk	74	0.26	0.44	0	1
Reasons for not implementing the training practices:					
Not used to following practices	74	0.12	0.33	0	1
Not enough time to follow practices	74	0.16	0.37	0	1
Too expensive to follow practices	74	0.04	0.20	0	1
Do not think practices are important	74	0.04	0.20	0	1

Notes: Data is from the Endline 1 survey conducted in September 2022. Questions on the perceptions of training were asked in a yes/no format. Vendors were asked why they found the training to be difficult and could select from multiple different options.

Table A19: Mediation Analysis

	(1) Food-safety practices (0-13)	(2) Profits, monthly	(3) Sales monthly	(4) Expend., monthly	(5) Custom., monthly
	M: Large equip. (0-4)	M: Overall quality (0-20)			
Mediated (%)	77.8	21.6	19.0	22.9	40.8
p-value	0.01	0.01	0.02	0.04	0.00

Notes: Each row reports the proportion of the total treatment effect mediated by the variable labeled "M." Column (1) uses "Large equipment" as the mediator and food safety practices as the outcome. Columns (2)–(5) use "Overall quality" as the mediator and business outcomes (monthly profits, sales, expenditures, and customer numbers) as outcomes. Mediation shares and p-values are estimated using the `mediate` command in Stata with Poisson models and standard errors clustered at the class level. All regressions follow the same specification as in the main analysis, including treatment assignment, strata fixed effects, the appropriate pre-treatment controls, and interaction terms between treatment and the mediator.

G Robustness Checks and Treatment Effects Heterogeneity

Table A20: Treatment Effects on Equipment Usage, Small Supplies and Food-Safety Practices (Periods 3 and 4 included)

	(1) Overall quality (0-20)	(2) Large equip. (0-4)	(3) Small supplies (0-3)	(4) Food-safety practices (0-13)	(5) Chlorine $\mathbb{1}(> 0.20)$ (ppm)
Equipment (T1)	0.296***	1.133***	0.304*	0.099***	0.887***
	(0.023)	(0.077)	(0.130)	(0.026)	(0.018)
	[0.001]	[0.001]	[0.017]	[0.001]	[0.001]
w/ training (T2)	0.262***	1.035***	0.236	0.092**	0.903***
	(0.028)	(0.087)	(0.135)	(0.029)	(0.016)
	[0.001]	[0.001]	[0.041]	[0.002]	[0.001]
Equipment (T1) x post	-0.055	-0.216*	-0.161	-0.029	-0.841***
	(0.031)	(0.090)	(0.150)	(0.040)	(0.022)
	[0.041]	[0.015]	[0.118]	[0.118]	[0.001]
Equipment (T2) x post	-0.060	-0.203*	-0.283	-0.028	-0.849***
	(0.035)	(0.091)	(0.166)	(0.046)	(0.025)
	[0.041]	[0.020]	[0.041]	[0.118]	[0.001]
Control mean:	7.17	0.88	0.29	6.01	0.04
Implied T1 effect (%):	34.5	210.5	35.6	10.4	2144.6
Implied T2 effect (%):	30.0	181.5	26.6	9.6	2184.2
Clusters:	108	108	108	108	106
Observations:	3587	3587	3587	3587	2517
p-value T1-T2:	0.13	0.10	0.60	0.80	0.27

Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when equipment was delivered) up until the second endline survey in February 2023. Outcome variables in Columns (1)-(4) are equal to the number of components of each count variable observed at the time of data collection. We use a Poisson regression model for estimation. Whereas, in Column (5) the outcome variable is a binary variable taking value one if the amount of chlorine is above 0.20 ppm, and zero otherwise. "post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. All regressions include strata fixed effects and the pre-treatment average sanitary equipment observed at the kiosk. To increase precision, we also include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A21: Treatment Effects on Equipment Usage, Small Supplies and Food-Safety Practices (OLS Estimates)

	(1) Overall quality (0-20)	(2) Large equip. (0-4)	(3) Small supplies (0-3)	(4) Food-safety practices (0-13)
Equipment (T1)	2.421***	1.747***	0.084*	0.591***
	(0.188)	(0.101)	(0.035)	(0.170)
	[0.001]	[0.001]	[0.027]	[0.002]
w/ training (T2)	2.162***	1.512***	0.045	0.605**
	(0.233)	(0.143)	(0.039)	(0.182)
	[0.001]	[0.001]	[0.114]	[0.003]
Equipment (T1) x post	-0.461	-0.297*	-0.005	-0.159
	(0.252)	(0.121)	(0.047)	(0.243)
	[0.054]	[0.024]	[0.298]	[0.159]
Equipment (T2) x post	-0.584*	-0.272*	-0.077	-0.235
	(0.279)	(0.126)	(0.045)	(0.277)
	[0.038]	[0.038]	[0.063]	[0.133]
Control mean:	7.17	0.88	0.29	6.01
Implied T1 effect (%):	33.7	199.1	29.3	9.8
Implied T2 effect (%):	30.1	172.3	15.7	10.1
Clusters:	108	108	108	108
Observations:	3587	3587	3587	3587
Adjusted R^2:	0.48	0.56	0.39	0.44
p-value T1-T2:	0.22	0.10	0.41	0.94

Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcome variables are linear indices calculated as the count of the observed "best behaviors" divided by the number of maximum "best behaviors" in that category. We exclude chlorine here as it is estimated with OLS in Table 7. "post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. All OLS regressions include strata fixed effects and the pre-treatment average sanitary equipment observed at the kiosk. To increase precision, we also include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors at the cluster level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Heterogeneity in treatment effects. We examine heterogeneous treatment effects by area, cluster size, and kiosk size, as well as distributional effects across the outcome distribution. First, for area, we distinguish between the three neighborhoods in which the study was conducted: Dalhousie, Hazra, and Sector V. While all three host numerous food vendors preparing meals on-site, they differ in clientele and urban context. Sector V is an affluent IT and business district with cleaner streets and relatively safer vendor practices. Dalhousie, the city's central business district, is densely populated with food-only vendors serving office workers. Hazra is more mixed, with lower average income levels and a broader range of customer types—including residents, patients, shoppers, and pilgrims—as well as a greater presence of non-food vendors. Second, for cluster size, we adopt a classification based on the number of sampled food vendors operating within a 30-meter radius. Specifically, we define isolated vendors (cluster size = 1) as the reference group, and compare them to vendors in medium clusters (2–3 vendors) and large clusters (4 or more vendors). This classification, grounded in our experimental design, captures variation in the degree of local vendor density and reflects increasing levels of market competition and potential peer effects. Third, we define kiosk size based on staffing: small kiosks are those operated by the owner alone or with one additional employee, while large kiosks have two or more employees. Given the minimal differences between treatment groups in the main specification (Table 7), we pool T1 and T2 for all heterogeneity analyses. Lastly, we complement these results by examining distributional treatment effects on quality-related outcomes, which provide further insights into the types of vendors most affected by the intervention and the extent of behavioral change.

By Area. As shown in Table A22, treatment effects on overall equipment usage, small supplies, and food safety practices are broadly similar across the three study areas. However, some mixed patterns emerge when disaggregating specific components. During the treatment period, vendors in Dalhousie and Sector V are significantly more likely than those in Hazra to have a professional-looking handwashing station (Table A23). These vendors also report better cleanliness practices: they are more likely to maintain clean garbage bins, less likely to have garbage on the ground, and more likely to use disposable plates (Tables A24 and A25). At the same time, however, they are less likely to have clean towels, suggesting that improvements were not uniform across all hygiene-related behaviors.

In the post-treatment period, some of these gains appear to reverse. Vendors in Dalhousie, in particular, are more likely than those in Hazra to exhibit a decline in visible cleanliness standards—specifically, they are more likely to have garbage on the ground and less likely to continue using disposable plates. This suggests that the withdrawal of the subsidy may have had a greater negative impact on sustained hygiene behaviors in areas where baseline compliance was initially higher. Vendors in Sector V, while better off in the treatment phase, also do not show persistently stronger effects once the intervention ends.

By Cluster Size. Table A26 indicates that treatment effects do not vary meaningfully across vendor clusters of different sizes—whether vendors operate alone, in small groups, or in larger groups. When breaking down the quality components (Table A27-A29), we find that vendors in medium and large clusters are significantly less likely to wear a clean apron during the treatment period, potentially reflecting free-riding or peer norm effects in more crowded vendor settings. In the post-treatment phase, vendors in large clusters are also less likely to maintain a professional drinking water facility, suggesting that sustained effort may be harder to coordinate or maintain in denser vendor environments.

By Kiosk Size. Treatment effects do not differ significantly between small kiosks and larger kiosks, as shown in Tables A30-A33. Both types of vendors exhibit comparable adoption rates for equipment, small supplies, and food safety practices during and after the intervention period. This suggests that capacity constraints related to labor size are not a key determinant of take-up or sustainability of food safety improvements in this context.

Distributional treatment effects. In addition to subgroup comparisons, we assess how the treatment shifted the distribution of hygiene-related behaviors. Figure A14-a) presents the average predicted probabilities for overall quality, measured as the number of observed best-practice indicators (ranging from zero to twenty). The treatment shifts the distribution rightward: the probability of observing fewer than 5 good practices drops sharply, while the probability of observing higher counts increases—suggesting broad-based improvements in quality. Figures A14-b), A14-c), and A14-d) disaggregate these effects into three subcomponents: large equipment, small supplies, and food-safety practices. In all three domains, the treatment reduces the likelihood of very low scores and increases the prevalence of higher scores. The shift is particularly visible in large equipment (Panel b), where the probability of observing no items falls steeply and the modal count increases to two. Effects on small supplies and food handling (Panels c and d) are consistent in direction but smaller in size. These patterns confirm that the intervention led to meaningful upgrades among vendors at the lower end of the quality distribution. However, consistent with earlier findings, the magnitude of change for behaviors requiring daily effort is visually small. Thus, while the program reduced the prevalence of very poor hygiene standards, sustaining high-frequency practices remains a challenge.

Table A22: Treatment Effects on Equipment Usage, Small Supplies and Food-Safety Practices By Area

	(1) Overall quality (0-20)	(2) Large equip. (0-4)	(3) Small supplies (0-3)	(4) Food-safety practices (0-13)	(5) Chlorine $\mathbb{1}(>0.20)$ (ppm)
Equipment (T1 or T2)	0.228***	1.058***	0.006	0.090	0.878***
	(0.046)	(0.149)	(0.304)	(0.048)	(0.030)
	[0.001]	[0.001]	[1.000]	[0.223]	[0.001]
Equipment (T1 or T2) x Dalhousie	0.062	0.026	0.293	0.014	0.001
	(0.052)	(0.167)	(0.337)	(0.054)	(0.035)
	[0.479]	[1.000]	[0.625]	[1.000]	[1.000]
Equipment (T1 or T2) x Sector V	0.084	0.047	0.336	0.014	0.060
	(0.076)	(0.223)	(0.404)	(0.078)	(0.039)
	[0.479]	[1.000]	[0.648]	[1.000]	[0.319]
Equipment (T1 or T2) x Post	-0.040	-0.275	-0.704**	-0.011	-0.860***
	(0.075)	(0.181)	(0.266)	(0.073)	(0.032)
	[0.929]	[0.319]	[0.045]	[1.000]	[0.001]
Equipment (T1 or T2) x Post x Dalhousie	0.005	0.100	0.528	-0.012	0.047
	(0.081)	(0.205)	(0.320)	(0.081)	(0.044)
	[1.000]	[0.939]	[0.310]	[1.000]	[0.489]
Equipment (T1 or T2) x Post x Sector V	-0.066	0.047	0.605	-0.038	-0.040
	(0.098)	(0.241)	(0.349)	(0.101)	(0.043)
	[0.857]	[1.000]	[0.295]	[1.000]	[0.576]
Control mean:	7.17	0.88	0.29	6.01	0.04
Clusters:	108	108	108	108	106
Observations:	3587	3587	3587	3587	2517

Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcome variables in Columns (1)-(4) are equal to the number of components of each count variable observed at the time of data collection. We use a Poisson regression model for estimation. Whereas, in Column (5) the outcome variable is a binary variable taking value one if the amount of chlorine is above 0.20 ppm, and zero otherwise. We use a Poisson regression model for estimation of Columns (1)-(4) and OLS for column (5). "Post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. Hazra is the leave-out Sector. We include the full set of interactions; only those related to the treatment effect are displayed here. All regressions include strata fixed effects and the pre-treatment average sanitary equipment observed at the kiosk. To increase precision, we also include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A23: Treatment Effects on Equipment Usage and Small Supplies By Index Component and By Area

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Large equipment				Small supplies		
	Handwash facility	Water container	Drinking water	Garbage bin	Clean apron	Hair cover	Handwash soap
Equipment (T1 or T2)	0.148*	0.160	0.369***	0.650***	0.011	0.039**	-0.102**
	(0.060)	(0.087)	(0.069)	(0.078)	(0.037)	(0.015)	(0.038)
	[0.073]	[0.177]	[0.001]	[0.001]	[0.727]	[0.053]	[0.053]
Equipment (T1 or T2) x Dalhousie	0.243**	0.064	-0.074	-0.011	0.063	-0.019	0.095*
	(0.087)	(0.094)	(0.084)	(0.089)	(0.043)	(0.015)	(0.046)
	[0.046]	[0.583]	[0.532]	[0.788]	[0.266]	[0.323]	[0.134]
Equipment (T1 or T2) x Sector V	0.600***	0.098	-0.003	0.028	0.096	0.006	0.044
	(0.095)	(0.120)	(0.091)	(0.110)	(0.064)	(0.029)	(0.058)
	[0.001]	[0.583]	[0.795]	[0.752]	[0.266]	[0.758]	[0.583]
Equipment (T1 or T2) x Post	0.101	0.014	-0.233**	-0.089	-0.169**	-0.039***	0.085
	(0.143)	(0.061)	(0.077)	(0.068)	(0.055)	(0.011)	(0.044)
	[0.583]	[0.758]	[0.024]	[0.323]	[0.024]	[0.005]	[0.150]
Equipment (T1 or T2) x Post x Dalhousie	0.107	-0.091	0.015	-0.004	0.117	0.030	-0.041
	(0.164)	(0.068)	(0.090)	(0.078)	(0.065)	(0.015)	(0.055)
	[0.586]	[0.314]	[0.782]	[0.795]	[0.181]	[0.150]	[0.583]
Equipment (T1 or T2) x Post x Sector V	-0.157	-0.193	-0.032	0.008	0.116	-0.002	-0.013
	(0.163)	(0.113)	(0.098)	(0.082)	(0.069)	(0.020)	(0.069)
	[0.490]	[0.214]	[0.722]	[0.790]	[0.214]	[0.790]	[0.772]
Control mean:	0.06	0.75	0.02	0.05	0.04	0.00	0.25
Clusters:	108	108	108	108	108	108	108
Observations:	3587	3577	3587	3587	3587	3587	3548
Adjusted R^2:	0.51	0.33	0.27	0.51	0.17	0.07	0.54

Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcomes are binary variables taking value one if "Best behaviour" in the described category is observed, and zero otherwise. "Post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. Hazra is the leave-out Sector. We include the full set of interactions; only those related to the treatment effect are displayed here. All OLS regressions include strata fixed effects and the pre-treatment average sanitary infrastructure observed at the kiosk. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors are always clustered at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A24: Treatment Effects on Food-Safety Practices By Index Component and By Area (Part 1)

	(1) Uses soap for dishes	(2) Dish water clean	(3) Garbage bin clean	(4) Garbage on ground	(5) Water on ground	(6) Food on ground
Equipment (T1 or T2)	-0.021	0.079	-0.088	0.254**	0.138	-0.057
	(0.036)	(0.042)	(0.102)	(0.076)	(0.071)	(0.069)
	[0.477]	[0.139]	[0.365]	[0.007]	[0.138]	[0.379]
Equipment (T1 or T2) x Dalhousie	0.053	-0.003	0.298**	-0.280***	-0.139	0.035
	(0.039)	(0.052)	(0.109)	(0.083)	(0.082)	(0.077)
	[0.249]	[0.730]	[0.040]	[0.007]	[0.172]	[0.518]
Equipment (T1 or T2) x Sector V	0.068	0.008	0.426***	-0.221*	-0.080	0.009
	(0.040)	(0.064)	(0.114)	(0.106)	(0.109)	(0.088)
	[0.172]	[0.684]	[0.004]	[0.112]	[0.416]	[0.692]
Equipment (T1 or T2) x Post	0.020	0.064	0.260	-0.230*	-0.094	0.085
	(0.034)	(0.061)	(0.135)	(0.091)	(0.084)	(0.106)
	[0.477]	[0.322]	[0.139]	[0.055]	[0.300]	[0.392]
Equipment (T1 or T2) x Post x Dalhousie	-0.019	-0.056	-0.209	0.237*	0.065	-0.056
	(0.038)	(0.074)	(0.139)	(0.101)	(0.096)	(0.114)
	[0.503]	[0.406]	[0.201]	[0.074]	[0.447]	[0.503]
Equipment (T1 or T2) x Post x Sector V	-0.074	-0.087	-0.279	0.189	0.021	-0.082
	(0.040)	(0.096)	(0.148)	(0.117)	(0.124)	(0.124)
	[0.142]	[0.358]	[0.139]	[0.189]	[0.664]	[0.452]
Control mean:	0.96	0.75	0.05	0.27	0.31	0.76
Clusters:	108	108	108	108	108	108
Observations:	2348	3390	2821	3555	3551	3540
Adjusted R^2:	0.10	0.35	0.43	0.36	0.39	0.48

Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcomes are binary variables taking value one if "Best behaviour" in the described category is observed, and zero otherwise. "Post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. Hazra is the leave-out Sector. We include the full set of interactions; only those related to the treatment effect are displayed here. All OLS regressions include strata fixed effects and the pre-treatment average sanitary infrastructure observed at the kiosk. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors are always clustered at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A25: Treatment Effects on Food-Safety Practices By Index Component and By Area (Part 2)

	(1) Disposable plates	(2) Clean towel	(3) Counter clean	(4) Cooked food covered	(5) Raw food covered	(6) Use tongs or spoons	(7) Wash hands with soap
Equipment (T1 or T2)	-0.109	0.141***	-0.000	0.075	0.131	-0.018	-0.060
	(0.065)	(0.034)	(0.048)	(0.087)	(0.078)	(0.028)	(0.036)
	[0.294]	[0.002]	[0.830]	[0.563]	[0.294]	[0.563]	[0.294]
Equipment (T1 or T2) x Dalhousie	0.242**	-0.147**	0.012	0.025	-0.032	0.012	0.084*
	(0.086)	(0.053)	(0.061)	(0.095)	(0.083)	(0.037)	(0.037)
	[0.058]	[0.058]	[0.754]	[0.718]	[0.712]	[0.712]	[0.167]
Equipment (T1 or T2) x Sector V	0.021	-0.103*	0.057	0.043	-0.096	-0.028	0.063
	(0.093)	(0.050)	(0.072)	(0.105)	(0.087)	(0.067)	(0.043)
	[0.732]	[0.200]	[0.563]	[0.707]	[0.477]	[0.707]	[0.368]
Equipment (T1 or T2) x Post	0.226*	-0.046	0.038	-0.205	-0.115	-0.189	0.044
	(0.106)	(0.048)	(0.071)	(0.122)	(0.096)	(0.097)	(0.043)
	[0.174]	[0.545]	[0.634]	[0.294]	[0.411]	[0.222]	[0.506]
Equipment (T1 or T2) x Post x Dalhousie	-0.337**	0.084	-0.005	0.149	0.044	0.166	-0.067
	(0.122)	(0.063)	(0.087)	(0.139)	(0.102)	(0.121)	(0.044)
	[0.058]	[0.368]	[0.797]	[0.495]	[0.707]	[0.368]	[0.368]
Equipment (T1 or T2) x Post x Sector V	-0.163	0.024	-0.030	0.053	0.151	0.349*	-0.064
	(0.113)	(0.068)	(0.089)	(0.154)	(0.112)	(0.135)	(0.047)
	[0.368]	[0.712]	[0.712]	[0.712]	[0.368]	[0.080]	[0.368]
Control mean:	0.46	0.45	0.76	0.58	0.17	0.83	0.05
Clusters:	108	108	108	108	108	108	108
Observations:	3587	3587	3587	3587	2991	3587	3505
Adjusted R^2:	0.43	0.63	0.36	0.29	0.46	0.41	0.10

Note: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcomes are binary variables taking value one if "Best behaviour" in the described category is observed, and zero otherwise. "Post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. Hazra is the leave-out Sector. We include the full set of interactions; only those related to the treatment effect are displayed here. All OLS regressions include strata fixed effects and the pre-treatment average sanitary infrastructure observed at the kiosk. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors are always clustered at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A26: Treatment Effects on Equipment Usage, Small Supplies and Food-Safety Practices By Cluster Size

	(1) Overall quality (0-20)	(2) Large equip. (0-4)	(3) Small supplies (0-3)	(4) Food-safety practices (0-13)	(5) Chlorine $\mathbb{1}(> 0.20)$ (ppm)
Equipment (T1 or T2)	0.385***	1.480***	0.561*	0.116	0.806***
	(0.110)	(0.299)	(0.250)	(0.111)	(0.058)
	[0.003]	[0.001]	[0.089]	[0.306]	[0.001]
Equipment (T1 or T2) x Medium Cluster	-0.128	-0.428	-0.269	-0.040	0.088
	(0.115)	(0.318)	(0.307)	(0.116)	(0.062)
	[0.300]	[0.233]	[0.392]	[0.409]	[0.233]
Equipment (T1 or T2) x Large Cluster	-0.101	-0.436	-0.447	-0.005	0.117
	(0.112)	(0.306)	(0.318)	(0.114)	(0.060)
	[0.380]	[0.233]	[0.233]	[0.506]	[0.157]
Equipment (T1 or T2) x Post	-0.168	-0.512	-0.206	-0.096	-0.717***
	(0.122)	(0.267)	(0.266)	(0.144)	(0.071)
	[0.233]	[0.157]	[0.405]	[0.405]	[0.001]
Equipment (T1 or T2) x Post x Medium Cluster	0.151	0.416	-0.088	0.093	-0.151
	(0.128)	(0.288)	(0.359)	(0.152)	(0.077)
	[0.280]	[0.233]	[0.458]	[0.405]	[0.157]
Equipment (T1 or T2) x Post x Large Cluster	0.100	0.245	0.163	0.067	-0.144
	(0.127)	(0.287)	(0.319)	(0.155)	(0.075)
	[0.405]	[0.395]	[0.405]	[0.405]	[0.159]
Control mean:	7.17	0.88	0.29	6.01	0.04
Clusters:	108	108	108	108	106
Observations:	3587	3587	3587	3587	2517

Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcome variables in Columns (1)-(4) are equal to the number of components of each count variable observed at the time of data collection. We use a Poisson regression model for estimation. Whereas, in Column (5) the outcome variable is a binary variable taking value one if the amount of chlorine is above 0.20 ppm, and zero otherwise. We use a Poisson regression model for estimation of Columns (1)-(4) and OLS for column (5). "Post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. "Medium Cluster" is a cluster size equal to 2 or 3 sampled vendors. "Large Cluster" is a cluster size equal to 3 or more sampled vendors. A cluster size of 1 is the leave-out. We include the full set of interactions; only those related to the treatment effect are displayed here. All regressions include strata fixed effects and the pre-treatment average sanitary equipment observed at the kiosk. To increase precision, we also include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A27: Treatment Effects on Equipment Usage and Small Supplies By Index Component and By Cluster Size

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Large equipment				Small supplies		
	Handwash facility	Water container	Drinking water	Garbage bin	Clean apron	Hair cover	Handwash soap
Equipment (T1 or T2)	0.520***	0.392***	0.342***	0.610***	0.251**	0.035	-0.040
	(0.092)	(0.105)	(0.074)	(0.116)	(0.075)	(0.018)	(0.053)
	[0.001]	[0.004]	[0.001]	[0.001]	[0.010]	[0.209]	[0.807]
Equipment (T1 or T2) x Medium Cluster	-0.085	-0.151	-0.101	0.057	-0.182*	-0.021	0.010
	(0.110)	(0.115)	(0.087)	(0.130)	(0.078)	(0.020)	(0.059)
	[0.807]	[0.575]	[0.643]	[1.000]	[0.093]	[0.705]	[1.000]
Equipment (T1 or T2) x Large Cluster	-0.088	-0.256*	0.093	0.031	-0.234**	0.011	0.006
	(0.124)	(0.111)	(0.098)	(0.123)	(0.079)	(0.018)	(0.061)
	[0.830]	[0.093]	[0.723]	[1.000]	[0.025]	[0.915]	[1.000]
Equipment (T1 or T2) x Post	0.023	-0.213	-0.046	-0.069	-0.147*	-0.007	0.033
	(0.156)	(0.113)	(0.065)	(0.075)	(0.064)	(0.026)	(0.065)
	[1.000]	[0.226]	[0.830]	[0.723]	[0.093]	[1.000]	[0.999]
Equipment (T1 or T2) x Post x Medium Cluster	0.160	0.139	-0.118	-0.031	0.092	0.000	-0.008
	(0.170)	(0.126)	(0.080)	(0.086)	(0.074)	(0.028)	(0.077)
	[0.723]	[0.643]	[0.478]	[1.000]	[0.621]	[1.000]	[1.000]
Equipment (T1 or T2) x Post x Large Cluster	0.046	0.164	-0.320***	0.002	0.079	-0.034	0.063
	(0.191)	(0.117)	(0.084)	(0.089)	(0.074)	(0.028)	(0.073)
	[1.000]	[0.509]	[0.003]	[1.000]	[0.643]	[0.643]	[0.738]
Control mean:	0.06	0.75	0.02	0.05	0.04	0.00	0.25
Clusters:	108	108	108	108	108	108	108
Observations:	3587	3577	3587	3587	3587	3587	3548
Adjusted R^2:	0.47	0.33	0.28	0.50	0.17	0.07	0.54

Notes: Data from monitoring surveys (random audits) measured during the entire study perio (including periods 3 and 4 when the equipment was being delivered)d up until the second endline survey in February 2023. Outcomes are binary variables taking value one if "Best behaviour" in the described category is observed, and zero otherwise. "Post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. "Medium Cluster" is a cluster size equal to 2 or 3 sampled vendors. "Large Cluster" is a cluster size equal to 3 or more sampled vendors. A cluster size of 1 is the leave-out. We include the full set of interactions; only those related to the treatment effect are displayed here. All OLS regressions include strata fixed effects and the pre-treatment average sanitary infrastructure observed at the kiosk. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors are always clustered at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A28: Treatment Effects on Food-Safety Practices By Index Component and By Cluster Size (Part 1)

	(1) Uses soap for dishes	(2) Dish water clean	(3) Garbage bin clean	(4) Garbage on ground	(5) Water on ground	(6) Food on ground
Equipment (T1 or T2)	0.058	0.111	0.132	0.041	-0.007	0.060
	(0.039)	(0.061)	(0.126)	(0.086)	(0.097)	(0.065)
	[0.566]	[0.375]	[1.000]	[1.000]	[1.000]	[1.000]
Equipment (T1 or T2) x Medium Cluster	-0.059	-0.049	0.068	-0.000	0.040	-0.122
	(0.044)	(0.072)	(0.140)	(0.098)	(0.108)	(0.080)
	[0.765]	[1.000]	[1.000]	[1.000]	[1.000]	[0.566]
Equipment (T1 or T2) x Large Cluster	0.001	-0.030	0.078	-0.004	0.069	-0.116
	(0.044)	(0.068)	(0.132)	(0.104)	(0.116)	(0.074)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[0.564]
Equipment (T1 or T2) x Post	-0.030	0.041	-0.058	-0.068	-0.057	-0.041
	(0.034)	(0.104)	(0.105)	(0.114)	(0.140)	(0.136)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Equipment (T1 or T2) x Post x Medium Cluster	0.030	-0.049	0.182	0.016	0.008	0.070
	(0.038)	(0.115)	(0.122)	(0.124)	(0.151)	(0.153)
	[1.000]	[1.000]	[0.566]	[1.000]	[1.000]	[1.000]
Equipment (T1 or T2) x Post x Large Cluster	-0.002	-0.011	0.094	0.032	-0.012	0.092
	(0.048)	(0.116)	(0.123)	(0.134)	(0.161)	(0.158)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Control mean:	0.96	0.75	0.05	0.27	0.31	0.76
Clusters:	108	108	108	108	108	108
Observations:	2348	3390	2821	3555	3551	3540
Adjusted R^2:	0.10	0.35	0.42	0.35	0.37	0.45

Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcomes are binary variables taking value one if "Best behaviour" in the described category is observed, and zero otherwise. "Post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. "Medium Cluster" is a cluster size equal to 2 or 3 sampled vendors. "Large Cluster" is a cluster size equal to 3 or more sampled vendors. A cluster size of 1 is the leave-out. We include the full set of interactions; only those related to the treatment effect are displayed here. All OLS regressions include strata fixed effects and the pre-treatment average sanitary infrastructure observed at the kiosk. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors are always clustered at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A29: Treatment Effects on Food-Safety Practices By Index Component and By Cluster Size (Part 2)

	(1) Disposable plates	(2) Clean towel	(3) Counter clean	(4) Cooked food covered	(5) Raw food covered	(6) Use tongs or spoons	(7) Wash hands with soap
Equipment (T1 or T2)	-0.079	0.064	0.136	-0.057	0.174**	-0.041	-0.028
	(0.104)	(0.065)	(0.085)	(0.083)	(0.065)	(0.074)	(0.029)
	[0.915]	[0.842]	[0.482]	[0.915]	[0.102]	[1.000]	[0.842]
Equipment (T1 or T2) x Medium Cluster	0.127	-0.076	-0.134	0.166	-0.092	0.023	0.033
	(0.118)	(0.073)	(0.096)	(0.097)	(0.074)	(0.080)	(0.036)
	[0.797]	[0.842]	[0.566]	[0.452]	[0.631]	[1.000]	[0.864]
Equipment (T1 or T2) x Large Cluster	0.116	0.012	-0.143	0.184*	-0.092	0.013	0.043
	(0.122)	(0.078)	(0.090)	(0.093)	(0.074)	(0.078)	(0.034)
	[0.842]	[1.000]	[0.484]	[0.380]	[0.631]	[1.000]	[0.631]
Equipment (T1 or T2) x Post	0.009	-0.018	-0.035	-0.149	-0.058	-0.019	0.029
	(0.069)	(0.097)	(0.131)	(0.151)	(0.090)	(0.110)	(0.025)
	[1.000]	[1.000]	[1.000]	[0.842]	[0.971]	[1.000]	[0.742]
Equipment (T1 or T2) x Post x Medium Cluster	0.062	0.061	0.051	0.022	-0.001	0.019	-0.041
	(0.091)	(0.104)	(0.142)	(0.168)	(0.105)	(0.137)	(0.031)
	[0.915]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[0.599]
Equipment (T1 or T2) x Post x Large Cluster	-0.104	-0.014	0.110	0.067	0.008	0.044	-0.055
	(0.108)	(0.106)	(0.142)	(0.171)	(0.101)	(0.135)	(0.032)
	[0.842]	[1.000]	[0.915]	[1.000]	[1.000]	[1.000]	[0.452]
Control mean:	0.46	0.45	0.76	0.58	0.17	0.83	0.05
Clusters:	108	108	108	108	108	108	108
Observations:	3587	3587	3587	3587	2991	3587	3505
Adjusted R^2:	0.42	0.63	0.36	0.29	0.46	0.41	0.10

Note: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcomes are binary variables taking value one if "Best behaviour" in the described category is observed, and zero otherwise. "Post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. "Medium Cluster" is a cluster size equal to 2 or 3 sampled vendors. "Large Cluster" is a cluster size equal to 3 or more sampled vendors. A cluster size of 1 is the leave-out. We include the full set of interactions; only those related to the treatment effect are displayed here. All OLS regressions include strata fixed effects and the pre-treatment average sanitary infrastructure observed at the kiosk. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors are always clustered at the cluster level. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A30: Treatment Effects on Equipment Usage, Small Supplies and Food-Safety Practices By Kiosk Size

	(1) Overall quality (0-20)	(2) Large equip. (0-4)	(3) Small supplies (0-3)	(4) Food-safety practices (0-13)	(5) Chlorine $\mathbb{1}(> 0.20)$ (ppm)
Equipment (T1 or T2)	0.247***	1.141***	0.426*	0.063	0.909***
	(0.036)	(0.123)	(0.216)	(0.038)	(0.025)
	[0.001]	[0.001]	[0.129]	[0.186]	[0.001]
Equipment (T1 or T2) x Large Kiosk	0.055	-0.072	-0.224	0.053	-0.023
	(0.035)	(0.123)	(0.239)	(0.040)	(0.035)
	[0.186]	[0.520]	[0.382]	[0.302]	[0.520]
Equipment (T1 or T2) x Post	-0.092	-0.303*	-0.403	-0.060	-0.870***
	(0.047)	(0.143)	(0.226)	(0.053)	(0.032)
	[0.129]	[0.115]	[0.174]	[0.372]	[0.001]
Equipment (T1 or T2) x Post x Large Kiosk	0.052	0.134	0.279	0.048	0.039
	(0.054)	(0.160)	(0.255)	(0.060)	(0.042)
	[0.382]	[0.418]	[0.377]	[0.418]	[0.382]
Control mean:	7.17	0.88	0.29	6.01	0.04
Clusters:	108	108	108	108	106
Observations:	3587	3587	3587	3587	2517

Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcome variables in Columns (1)-(4) are equal to the number of components of each count variable observed at the time of data collection. We use a Poisson regression model for estimation. Whereas, in Column (5) the outcome variable is a binary variable taking value one if the amount of chlorine is above 0.20 ppm, and zero otherwise. We use a Poisson regression model for estimation of Columns (1)-(4) and OLS for column (5). "Post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. "Large Kiosk" is equal to 1 if a kiosk at baseline as two or more employees (excluding the owner). Kiosks with one or no employees (excluding the owner) are the leave-out. We include the full set of interactions; only those related to the treatment effect are displayed here. All regressions include strata fixed effects and the pre-treatment average sanitary equipment observed at the kiosk. To increase precision, we also include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A31: Treatment Effects on Equipment Usage and Small Supplies By Index Component and By Kiosk Size

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Large equipment				Small supplies		
	Handwash facility	Water container	Drinking water	Garbage bin	Clean apron	Hair cover	Handwash soap
Equipment (T1 or T2)	0.402***	0.232***	0.295***	0.589***	0.041	0.030***	0.001
	(0.060)	(0.054)	(0.045)	(0.056)	(0.031)	(0.008)	(0.028)
	[0.001]	[0.001]	[0.001]	[0.001]	[0.412]	[0.002]	[0.868]
Equipment (T1 or T2) x Large Kiosk	0.066	-0.016	0.060	0.101	0.049	0.001	-0.059
	(0.062)	(0.055)	(0.048)	(0.065)	(0.039)	(0.012)	(0.032)
	[0.462]	[0.764]	[0.412]	[0.317]	[0.412]	[0.855]	[0.176]
Equipment (T1 or T2) x Post	0.074	-0.088	-0.254***	-0.116**	-0.050*	-0.031**	0.023
	(0.075)	(0.062)	(0.053)	(0.044)	(0.023)	(0.010)	(0.040)
	[0.483]	[0.387]	[0.001]	[0.028]	[0.079]	[0.012]	[0.764]
Equipment (T1 or T2) x Post x Large Kiosk	0.075	-0.001	0.027	0.047	-0.034	0.013	0.058
	(0.094)	(0.067)	(0.060)	(0.058)	(0.040)	(0.014)	(0.047)
	[0.589]	[0.868]	[0.764]	[0.584]	[0.569]	[0.532]	[0.412]
Control mean:	0.06	0.75	0.02	0.05	0.04	0.00	0.25
Clusters:	108	108	108	108	108	108	108
Observations:	3587	3577	3587	3587	3587	3587	3548
Adjusted R^2:	0.47	0.32	0.26	0.51	0.16	0.07	0.54

Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcomes are binary variables taking value one if "Best behaviour" in the described category is observed, and zero otherwise. "Post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. "Large Kiosk" is equal to 1 if a kiosk at baseline as two or more employees (excluding the owner). Kiosks with one or no employees (excluding the owner) are the leave-out. We include the full set of interactions; only those related to the treatment effect are displayed here. All OLS regressions include strata fixed effects and the pre-treatment average sanitary infrastructure observed at the kiosk. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors are always clustered at the cluster level. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A32: Treatment Effects on Food-Safety Practices By Index Component and By Kiosk Size (Part 1)

	(1) Uses soap for dishes	(2) Dish water clean	(3) Garbage bin clean	(4) Garbage on ground	(5) Water on ground	(6) Food on ground
Equipment (T1 or T2)	0.035	0.031	0.179**	0.105*	0.067	-0.083*
	(0.023)	(0.036)	(0.065)	(0.053)	(0.052)	(0.036)
	[0.364]	[0.913]	[0.054]	[0.206]	[0.492]	[0.134]
Equipment (T1 or T2) x Large Kiosk	-0.012	0.072	0.027	-0.105	-0.044	0.061
	(0.029)	(0.043)	(0.062)	(0.057)	(0.063)	(0.041)
	[0.913]	[0.301]	[0.913]	[0.231]	[0.913]	[0.380]
Equipment (T1 or T2) x Post	-0.022	0.031	0.055	-0.139*	-0.095	-0.004
	(0.022)	(0.054)	(0.079)	(0.058)	(0.075)	(0.067)
	[0.695]	[0.913]	[0.913]	[0.134]	[0.492]	[1.000]
Equipment (T1 or T2) x Post x Large Kiosk	0.015	-0.031	0.018	0.140*	0.062	0.053
	(0.026)	(0.063)	(0.083)	(0.068)	(0.087)	(0.084)
	[0.913]	[0.913]	[1.000]	[0.187]	[0.913]	[0.913]
Control mean:	0.96	0.75	0.05	0.27	0.31	0.76
Clusters:	108	108	108	108	108	108
Observations:	2348	3390	2821	3555	3551	3540
Adjusted R^2:	0.10	0.35	0.42	0.36	0.38	0.45

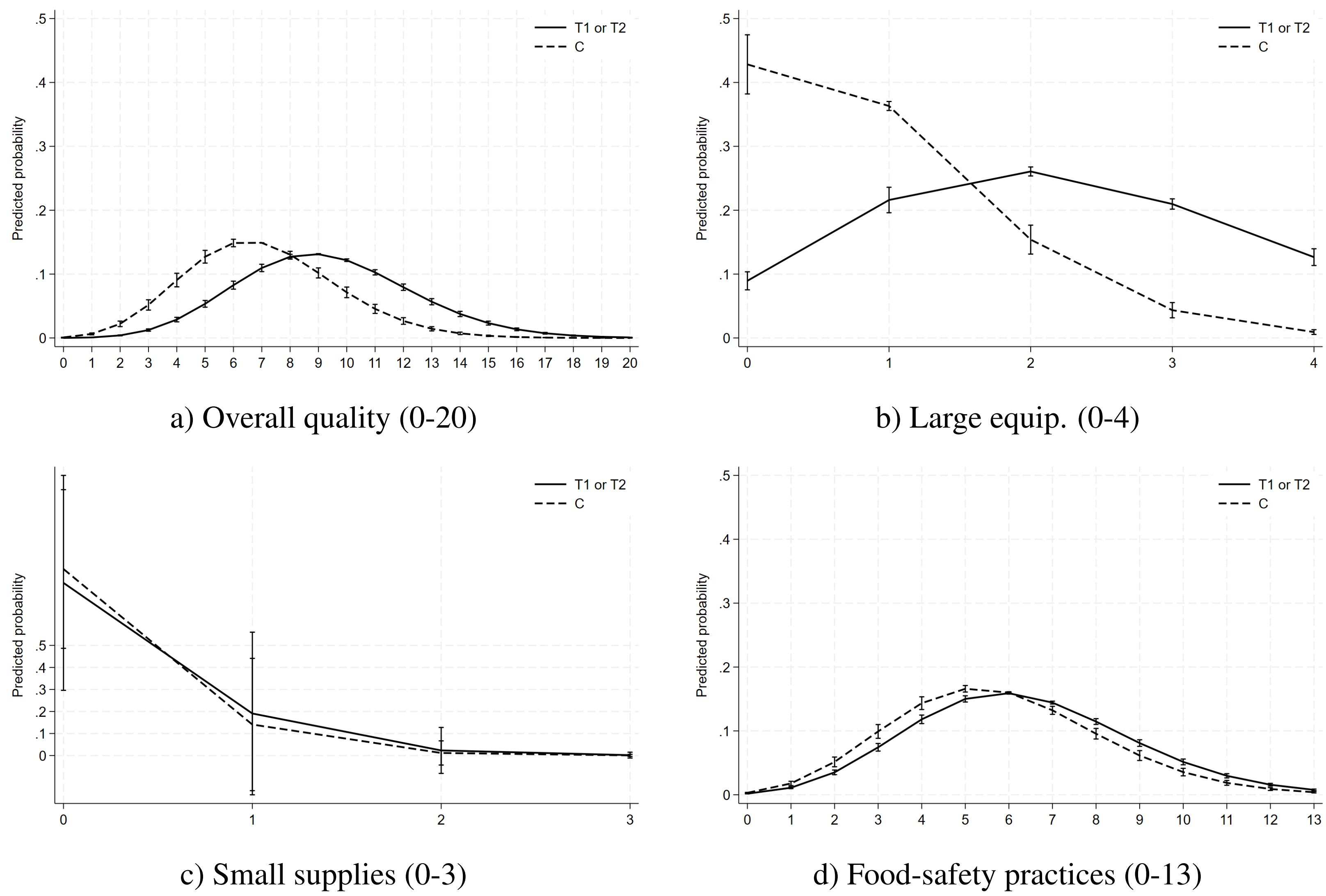
Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcomes are binary variables taking value one if "Best behaviour" in the described category is observed, and zero otherwise. "Post" is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. "Large Kiosk" is equal to 1 if a kiosk at baseline as two or more employees (excluding the owner). Kiosks with one or no employees (excluding the owner) are the leave-out. We include the full set of interactions; only those related to the treatment effect are displayed here. All OLS regressions include strata fixed effects and the pre-treatment average sanitary infrastructure observed at the kiosk. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors are always clustered at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A33: Treatment Effects on Food-Safety Practices By Index Component and By Kiosk Size (Part 2)

	(1) Disposable plates	(2) Clean towel	(3) Counter clean	(4) Cooked food covered	(5) Raw food covered	(6) Use tongs or spoons	(7) Wash hands with soap
Equipment (T1 or T2)	0.002	0.030	0.034	0.030	0.075	-0.055	-0.004
	(0.064)	(0.033)	(0.035)	(0.050)	(0.044)	(0.037)	(0.018)
	[1.000]	[1.000]	[1.000]	[1.000]	[0.800]	[0.970]	[1.000]
Equipment (T1 or T2) x Large Kiosk	0.043	0.003	-0.027	0.105	0.028	0.055	0.014
	(0.079)	(0.037)	(0.038)	(0.061)	(0.053)	(0.040)	(0.025)
	[1.000]	[1.000]	[1.000]	[0.800]	[1.000]	[1.000]	[1.000]
Equipment (T1 or T2) x Post	-0.045	-0.039	-0.019	-0.055	-0.051	0.038	-0.008
	(0.080)	(0.037)	(0.046)	(0.088)	(0.047)	(0.076)	(0.019)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Equipment (T1 or T2) x Post x Large Kiosk	0.061	0.071	0.074	-0.084	-0.008	-0.064	-0.007
	(0.089)	(0.047)	(0.061)	(0.103)	(0.058)	(0.089)	(0.027)
	[1.000]	[0.970]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Control mean:	0.46	0.45	0.76	0.58	0.17	0.83	0.05
Clusters:	108	108	108	108	108	108	108
Observations:	3587	3587	3587	3587	2991	3587	3505
Adjusted R^2:	0.42	0.63	0.35	0.29	0.46	0.41	0.10

Note: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcomes are binary variables taking value one if “Best behaviour” in the described category is observed, and zero otherwise. “Post” is a binary variable taking value one for observations after the end of the treatment period (after Endline 1), and zero otherwise. “Large Kiosk” is equal to 1 if a kiosk at baseline as two or more employees (excluding the owner). Kiosks with one or no employees (excluding the owner) are the leave-out. We include the full set of interactions; only those related to the treatment effect are displayed here. All OLS regressions include strata fixed effects and the pre-treatment average sanitary infrastructure observed at the kiosk. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, and years of experience. The results do not change with or without these controls. Standard errors are always clustered at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Figure A14: Distributional Treatment Effects on Equipment Usage, Small Supplies and Food-Safety Practices



Notes: Data from monitoring surveys (random audits). All figures are based on the estimates reported in Table 7. Panel a) displays the average predicted probabilities of overall quality, measured as the number of positive indicators (ranging from zero to twenty), for both treatment and control vendors. Panels b), c), and d) show the predicted probabilities for large equipment usage (0–4 items), small supplies (0–3 items), and food-safety practices (0–13 items), respectively.

H Robustness Checks for Business Outcomes

Table A34: Treatment Effects on Other Business Outcomes

	(1) Profits, expected	(2) Days work, expected	(3) Business assets PCA index	(4) Any savings	(5) Savings, per month
Equipment (T1 or T2)	0.033	0.003	0.374	-0.005	0.100
	(0.025)	(0.006)	(0.240)	(0.036)	(0.144)
	[0.888]	[0.896]	[0.888]	[1.000]	[0.896]
Control mean:	6.60	1.81	0.05	0.56	7.81
T effect (%):	3.3	0.3	45.4	-0.5	10.5
Clusters:	108	108	106	108	60
Observations:	2409	2419	254	522	145
Adjusted R^2:	0.54	0.34	0.63	0.50	0.47

Notes: Columns (1) and (2) use from monitoring surveys (random audits) and Columns (3)-(5) use data from the two Endline surveys. "Profits, expected" refers to the daily profit expections for the following week. "Days work, expected" refers to the number of days of work expected for the following week. "Business assets" are constructed using a large set of assets measured at both baseline and at Endline 1, and includes cookware, items for serving, furniture for customers, and storage containers. We aggregate this information using principal components analysis (PCA). "Any savings" is a variable equal to one if the vendor reports having any savings, zero otherwise. "Savings, per month" is the amount of monthly savings in logged terms that vendors report. OLS regressions are conducted at the vendor level and pooled for the entire study period (including periods 3 and 4 when the equipment was being delivered). "Equipment (T1 or T2)" equals one if the vendor belongs to T1 or T2, and zero otherwise. All regressions include strata fixed effects and logged average pre-treatment outcome measured at baseline. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, years of experience, as well as a control for whether the vendors keep their accounting. The results do not change with or without these controls. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A35: Treatment Effects on Business Outcomes (Periods 3 and 4 Included)

	(1)	(2)	(3)	(4)	(5)
	Panel A: Business Outcomes				
	Profits, monthly	Sales, monthly	Expend., monthly	Custom., monthly	Prices
Equipment (T1 or T2)	0.052*	0.068**	0.082**	0.068*	-0.001
	(0.022)	(0.026)	(0.029)	(0.031)	(0.014)
	[0.061]	[0.051]	[0.051]	[0.061]	[0.448]
Control mean:	9.65	11.33	11.11	7.45	3.68
T effect (%):	5.3	7.0	8.6	7.0	-0.1
Clusters:	108	108	108	108	103
Observations:	3015	3018	3018	2819	1576
Adjusted R^2:	0.53	0.70	0.68	0.55	0.98
	(1)	(2)	(3)	(4)	(5)
	Panel B: Labor Supply				
	Days, weekly	Total, daily hrs.	Prepare, daily hrs.	Sell, daily hrs.	Cleaning daily hrs.
Equipment (T1 or T2)	0.002	0.012	0.018	0.003	0.019
	(0.006)	(0.009)	(0.011)	(0.013)	(0.014)
	[0.448]	[0.228]	[0.160]	[0.448]	[0.189]
Control mean:	1.80	2.52	1.30	1.96	0.31
T effect (%):	0.2	1.2	1.8	0.3	2.0
Clusters:	108	108	108	108	108
Observations:	3018	3023	3023	3023	3023
Adjusted R^2:	0.19	0.46	0.24	0.49	0.47

Notes: Data from monitoring surveys (random audits), except for price data, which comes exclusively from Endline 2. All outcome variables are expressed in logs. Profits, sales, expenditures, and prices are reported in rupees. "Profits, monthly", "Sales, monthly," "Expenditures, monthly," and "Customers, monthly" are computed by multiplying daily values of the variable by the number of days worked in the previous week, then multiplying by four. “Prices” refers to the price of each item sold at a kiosk. “Days, weekly” refers to the number of days the kiosk was open in the previous week. “Total, daily hrs.” refers to the average number of hours per day the kiosk was open. “Prepare, daily hrs.” refers to the average number of hours per day the vendor spent preparing food. “Sell, daily hrs.” refers to average daily hours spent selling. “Cleaning, daily hrs.” refers to the average number of hours per day the vendor spent cleaning the kiosk. OLS regressions are conducted at the vendor level and pooled for the entire study period (including periods 3 and 4 when the equipment was being delivered), except for price regressions which are conducted at the item level and only for Endline 2. “Equipment (T1 or T2)” equals one if the vendor belongs to T1 or T2, and zero otherwise. All regressions include strata fixed effects and logged average pre-treatment outcome measured at baseline. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, years of experience, as well as a control for whether the vendors keep their accounting. The results do not change with or without these controls. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A36: Treatment Effects on Business Outcomes (Including Top 2%)

	(1)	(2)	(3)	(4)	(5)
			Panel A: Business Outcomes		
	Profits, monthly	Sales, monthly	Expend., monthly	Custom., monthly	Prices
Equipment (T1 or T2)	0.051*	0.072**	0.086**	0.061*	-0.001
	(0.025)	(0.026)	(0.029)	(0.031)	(0.014)
	[0.111]	[0.038]	[0.038]	[0.111]	[0.556]
Control mean:	9.67	11.34	11.12	7.46	3.68
T effect (%):	5.2	7.4	8.9	6.3	-0.1
Clusters:	108	108	108	108	103
Observations:	3102	3105	3105	2892	1576
Adjusted R^2:	0.51	0.71	0.68	0.55	0.98

	(1)	(2)	(3)	(4)	(5)
			Panel B: Labor Supply		
	Days, weekly	Total, daily hrs.	Prepare, daily hrs.	Sell, daily hrs.	Cleaning daily hrs.
Equipment (T1 or T2)	0.003	0.011	0.015	0.003	0.022
	(0.006)	(0.009)	(0.011)	(0.013)	(0.015)
	[0.458]	[0.273]	[0.218]	[0.556]	[0.184]
Control mean:	1.80	2.52	1.30	1.97	0.30
T effect (%):	0.3	1.1	1.5	0.3	2.3
Clusters:	108	108	108	108	108
Observations:	3105	3110	3110	3110	3110
Adjusted R^2:	0.19	0.47	0.24	0.50	0.44

Notes: Data from monitoring surveys (random audits), except for price data, which comes exclusively from Endline 2. All outcome variables are expressed in logs. Profits, sales, expenditures, and prices are reported in rupees. "Profits, monthly", "Sales, monthly," "Expenditures, monthly," and "Customers, monthly" are computed by multiplying daily values of the variable by the number of days worked in the previous week, then multiplying by four. "Prices" refers to the price of each item sold at a kiosk. "Days, weekly" refers to the number of days the kiosk was open in the previous week. "Total, daily hrs." refers to the average number of hours per day the kiosk was open. "Prepare, daily hrs." refers to the average number of hours per day the vendor spent preparing food. "Sell, daily hrs." refers to average daily hours spent selling. "Cleaning, daily hrs." refers to the average number of hours per day the vendor spent cleaning the kiosk. OLS regressions are conducted at the vendor level and pooled for the entire study period (including periods 3 and 4 when the equipment was being delivered), except for price regressions which are conducted at the item level and only for Endline 2. "Equipment (T1 or T2)" equals one if the vendor belongs to T1 or T2, and zero otherwise. All regressions include strata fixed effects and average logged pre-treatment outcome measured at baseline. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, years of experience, as well as a control for whether the vendors keep their accounting. The results do not change with or without these controls. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A37: Treatment Effects on Daily Business Outcomes

	(1) Profits, daily	(2) Sales, daily	(3) Expend., daily	(4) Custom., daily
Equipment (T1 or T2)	0.050*	0.063*	0.077**	0.065*
	(0.021)	(0.025)	(0.029)	(0.029)
	[0.029]	[0.029]	[0.029]	[0.029]
Control mean:	6.46	8.14	7.92	4.26
T effect (%):	5.1	6.5	8.0	6.7
Clusters:	108	108	108	108
Observations:	3019	3022	3022	2823
Adjusted R^2:	0.53	0.71	0.68	0.54

Notes: Data from monitoring surveys (random audits). All outcome are logged. Profits, sales, expenditures, and prices are average daily amounts for the previous week and in rupees. "Customers, daily" refers to the average number of customers per day in the previous week. OLS regressions are conducted at the vendor level and pooled for the entire study period (including periods 3 and 4 when the equipment was being delivered). "Equipment (T1 or T2)" equals one if the vendor belongs to T1 or T2, and zero otherwise. All regressions include strata fixed effects and average logged pre-treatment outcome measured at baseline. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, years of experience, as well as a control for whether the vendors keep their accounting. The results do not change with or without these controls. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A38: Treatment Effects on Business Outcomes (Linear Outcomes)

	(1)	(2)	(3)	(4)	(5)
			Panel A: Business Outcomes		
	Profits, monthly	Sales, monthly	Expend., monthly	Custom., monthly	Prices
Equipment (T1 or T2)	822*	6207**	6239**	100	-1
	(404)	(2222)	(2015)	(57)	(1)
	[0.134]	[0.029]	[0.026]	[0.173]	[0.262]
Control mean:	16544.33	92055.16	75530.29	1877.70	49.74
Clusters:	108	108	108	108	103
Observations:	3029	3029	3029	2828	1576
Adjusted R^2:	0.49	0.71	0.71	0.57	0.97
	(1)	(2)	(3)	(4)	(5)
			Panel B: Labor Supply		
	Days, weekly	Total, daily hrs.	Prepare, daily hrs.	Sell, daily hrs.	Cleaning daily hrs.
Equipment (T1 or T2)	-0.003	0.111	0.066	0.015	0.022
	(0.036)	(0.118)	(0.040)	(0.099)	(0.020)
	[0.659]	[0.279]	[0.173]	[0.659]	[0.262]
Control mean:	6.08	12.59	3.73	7.28	1.42
Clusters:	108	108	108	108	108
Observations:	3029	3029	3029	3029	3029
Adjusted R^2:	0.16	0.42	0.23	0.48	0.50

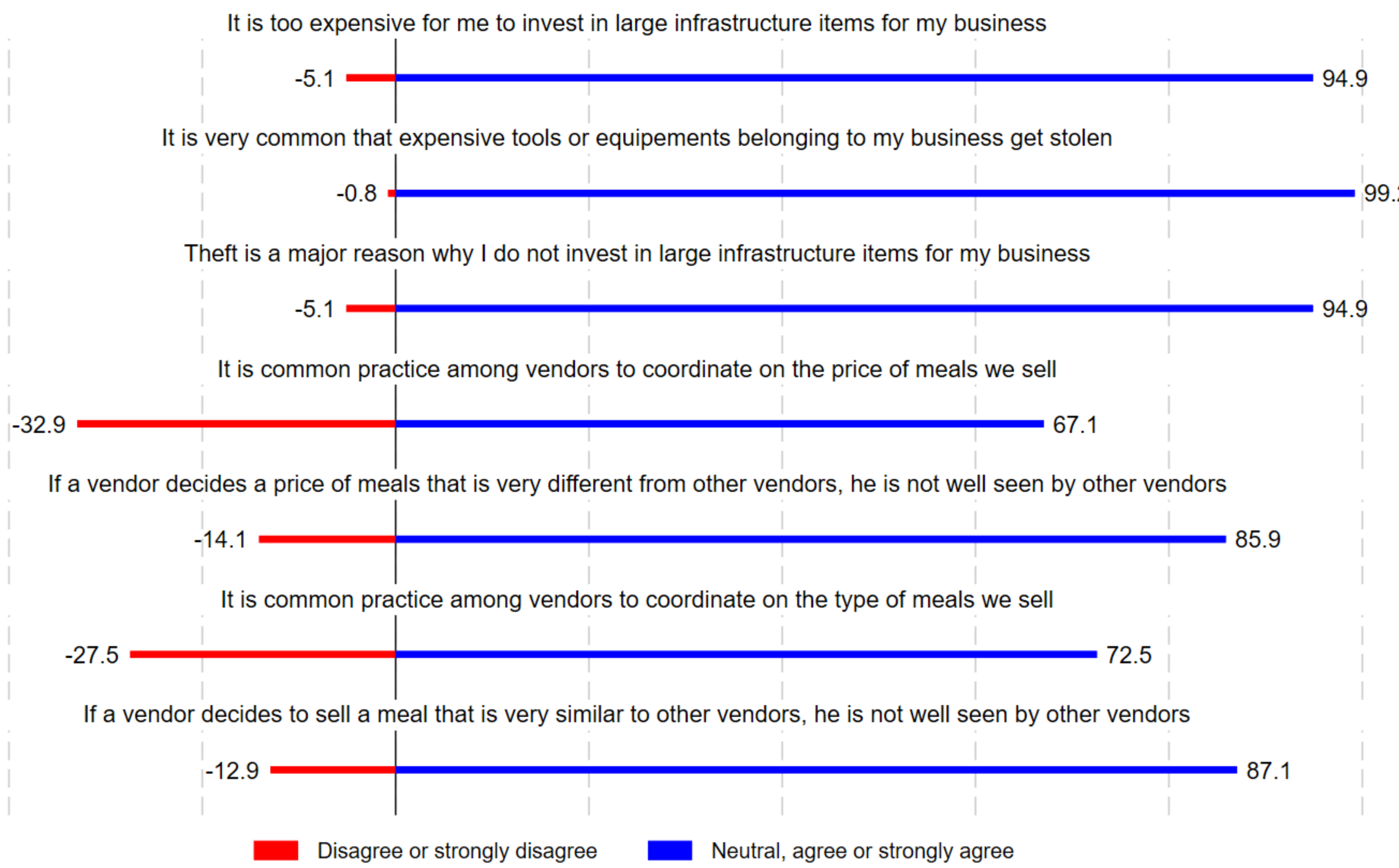
Notes: Data from monitoring surveys (random audits), except for price data, which comes exclusively from Endline 2. All outcome variables are expressed in logs. Profits, sales, expenditures, and prices are reported in rupees. "Profits, monthly", "Sales, monthly," "Expenditures, monthly," and "Customers, monthly" are computed by multiplying daily values of the variable by the number of days worked in the previous week, then multiplying by four. “Prices” refers to the price of each item sold at a kiosk. “Days, weekly” refers to the number of days the kiosk was open in the previous week. “Total, daily hrs.” refers to the average number of hours per day the kiosk was open. “Prepare, daily hrs.” refers to the average number of hours per day the vendor spent preparing food. “Sell, daily hrs.” refers to average daily hours spent selling. “Cleaning, daily hrs.” refers to the average number of hours per day the vendor spent cleaning the kiosk. OLS regressions are conducted at the vendor level and pooled for the entire study period (including periods 3 and 4 when the equipment was being delivered), except for price regressions which are conducted at the item level and only for Endline 2. “Equipment (T1 or T2)” equals one if the vendor belongs to T1 or T2, and zero otherwise. All regressions include strata fixed effects and average pre-treatment outcome measured at baseline. To increase precision, we also include a set of controls that predict the outcome variables. These include fixed effects for the survey period, interviewer, number of employees, years of experience, as well as a control for whether the vendors keep their accounting. The results do not change with or without these controls. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A39: Treatment Effects on Business Outcomes (Lasso)

	(1)	(2)	(3)	(4)	(5)
	Panel A: Business Outcomes				
	Profits, monthly	Sales, monthly	Expend., monthly	Custom., monthly	Prices
Equipment (T1 or T2)	0.059**	0.047	0.046	0.043	0.002
	(0.025)	(0.031)	(0.034)	(0.030)	(0.014)
	[0.089]	[0.210]	[0.210]	[0.210]	[0.277]
Control mean:	9.67	11.35	11.13	7.45	3.68
Clusters:	108	108	108	108	103
Observations:	3029	3029	3029	2828	1576
Adjusted R^2:	0.703	0.857	0.840	0.737	0.98
	(1)	(2)	(3)	(4)	(5)
	Panel B: Labor Supply				
	Days, weekly	Total, daily hrs.	Prepare, daily hrs.	Sell, daily hrs.	Cleaning daily hrs.
Equipment (T1 or T2)	0.003	-0.002	0.002	-0.008	-0.008
	(0.006)	(0.011)	(0.014)	(0.016)	(0.018)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Control mean:	1.80	2.51	1.30	1.95	0.30
Clusters:	108	108	108	108	108
Observations:	3029	3029	3029	3029	3029
Adjusted R^2:	0.278	0.620	0.444	0.617	0.565

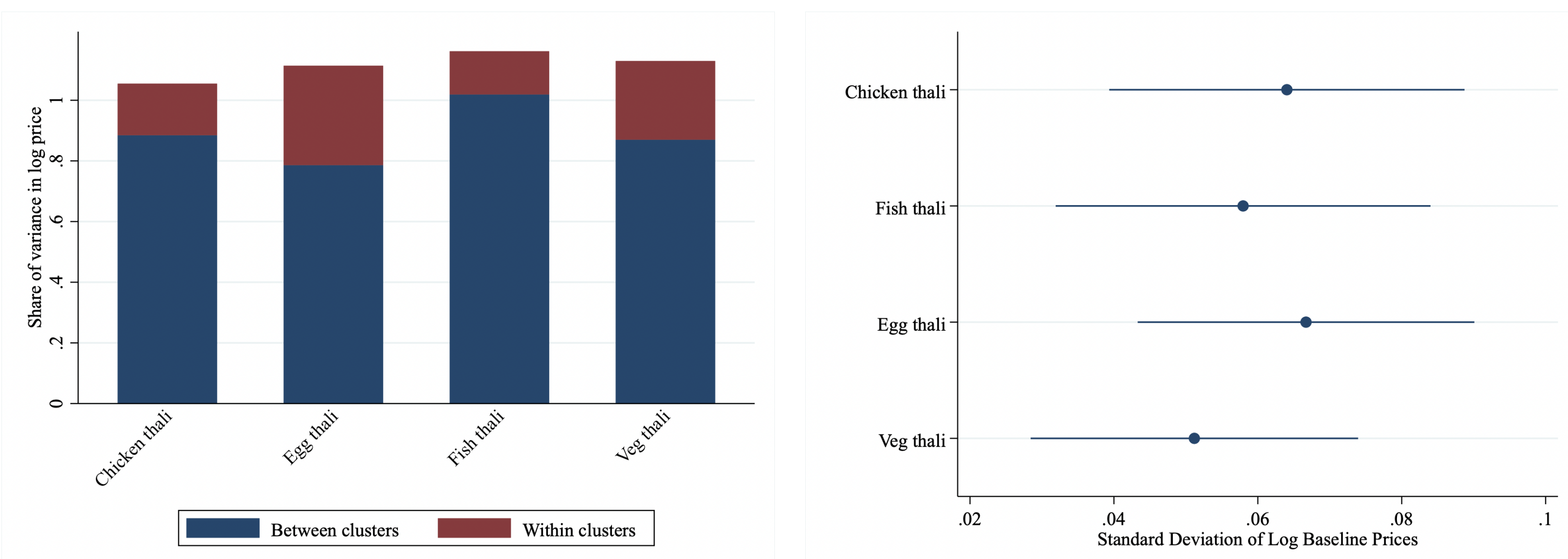
Notes: Data from monitoring surveys (random audits), except for price data, which comes exclusively from Endline 2. All outcome variables are expressed in logs. Profits, sales, expenditures, and prices are reported in rupees. "Profits, monthly", "Sales, monthly," "Expenditures, monthly," and "Customers, monthly" are computed by multiplying daily values of the variable by the number of days worked in the previous week, then multiplying by four. “Prices” refers to the price of each item sold at a kiosk. “Days, weekly” refers to the number of days the kiosk was open in the previous week. “Total, daily hrs.” refers to the average number of hours per day the kiosk was open. “Prepare, daily hrs.” refers to the average number of hours per day the vendor spent preparing food. “Sell, daily hrs.” refers to average daily hours spent selling. “Cleaning, daily hrs.” refers to the average number of hours per day the vendor spent cleaning the kiosk. All OLS regressions, except for prices, are conducted at the vendor level and pooled across all periods from period 5 (when the large equipment had been delivered to all vendors) up to the second endline survey in February 2023. We also trim the top 2% of reported profits to mitigate the influence of potential outliers. Whereas, the price regression in Column (5) is conducted at the item level and includes controls for baseline prices. “Equipment (T1 or T2)” is an indicator equal to one if the vendor belongs to treatment groups T1 or T2, and zero otherwise. Estimates are obtained via post-double selection lasso (Belloni et al., 2013). Strata fixed effects, period fixed effects, and interviewer fixed effects are always included. Additional pre-treatment covariates are chosen by the lasso from a rich baseline set. Standard errors are clustered at the vendor cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Figure A15: Rate how much you agree/disagree with each statement below



Notes: Data from Endline 2. Each question is asked using a Likert scale from 1 to 5, with 1 being strongly disagree and 5 being strongly agree.

Figure A16: Variation in prices between and within clusters



a) Between and within cluster variation

b) Variation in prices (%)

Notes: Data is from the baseline survey, where we collected a census of all items and their prices from vendors. Figure a) shows the decomposition of total variance into between and within cluster variation. Figure b) plots the difference in price standard deviation within clusters by item. Thali represents 26% of all items sold and is the relatively homogeneous across kiosks.

I Further Results on Spillover Effects.

Table A40: Treatment Effects on Small Supplies and Food-Safety Practices with Spillover Effects (Mean-Centered)

	(1) Small supplies (0-3)	(2) Food-safety practices (0-13)	(3) Chlorine $\mathbb{1}(>0.20)$ (ppm)
Equipment (T1 or T2)	0.170*	0.084***	0.461***
	(0.080)	(0.020)	(0.010)
	[0.168]	[0.001]	[0.001]
Vendors 0-400m	-0.004	-0.000	0.000
	(0.011)	(0.002)	(0.001)
	[1.000]	[1.000]	[1.000]
Vendors 400-800m	-0.012	-0.002	-0.002
	(0.014)	(0.003)	(0.001)
	[0.856]	[1.000]	[0.385]
T1 or T2 vendors 0-400m	0.022	-0.000	0.002
	(0.018)	(0.004)	(0.002)
	[0.610]	[1.000]	[0.856]
T1 or T2 vendors 400-800m	0.019	0.002	0.004
	(0.022)	(0.004)	(0.002)
	[0.856]	[1.000]	[0.269]
Control mean:	0.29	6.01	0.96
Clusters:	108	108	106
Observations:	3587	3587	2517

Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcome variable definitions and estimation details can be found in Table 7. We aggregate treatment groups for convenience. "Vendors 0-400m" is calculated as the number of vendors within a 0 to 400m radius centered around the mean; "Vendors 400-800m" is defined similarly. All regressions include strata fixed effects and the pre-treatment average sanitary equipment observed at the kiosk. To increase precision, we also include fixed effects for the survey period, interviewer, number of employees, and years of experience. Standard errors at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A41: Treatment Effects on Expenditures and Labor Supply (Mean-Centered)

	(1) Expend., monthly	(2) Days, weekly	(3) Total, daily hrs.	(4) Prepare, daily hrs.	(5) Sell, daily hrs.	(6) Cleaning daily hrs.
Equipment (T1 or T2)	0.085**	0.000	0.012	0.017	0.004	0.025
	(0.030)	(0.005)	(0.009)	(0.012)	(0.012)	(0.015)
	[0.031]	[0.378]	[0.184]	[0.171]	[0.378]	[0.147]
Vendors 0-400m	0.005	0.001	0.002	0.001	0.002	0.003*
	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	[0.184]	[0.174]	[0.142]	[0.276]	[0.221]	[0.076]
Vendors 400-800m	0.007	0.003***	0.003	0.001	0.004*	0.002
	(0.004)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
	[0.168]	[0.003]	[0.100]	[0.347]	[0.078]	[0.268]
T1 or T2 vendors 0-400m	-0.011*	-0.003**	-0.006**	-0.003	-0.007**	-0.006*
	(0.005)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
	[0.095]	[0.031]	[0.013]	[0.168]	[0.031]	[0.059]
T1 or T2 vendors 400-800m	-0.010	-0.005***	-0.005*	-0.001	-0.008**	-0.003
	(0.007)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)
	[0.184]	[0.001]	[0.059]	[0.378]	[0.031]	[0.300]
Control mean:	11.12	1.80	2.52	1.30	1.97	0.30
Clusters:	108	108	108	108	108	108
Observations:	3105	3105	3110	3110	3110	3110

Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcome variable definitions and estimation details can be found in Table 8. We aggregate treatment groups for convenience. “Vendors 0-400m” is calculated as the number of vendors within a 0 to 400m radius centered around the mean; “Vendors 400-800m” is defined similarly. All regressions include strata fixed effects and the pre-treatment average sanitary equipment observed at the kiosk. To increase precision, we also include fixed effects for the survey period, interviewer, number of employees, and years of experience. Standard errors at the cluster level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A42: Treatment Effects on Quality and Business Outcomes with Spillover Effects (Not Mean-Centered)

	(1) Overall quality (0-20)	(2) Large equip. (0-4)	(3) Profits, daily	(4) Sales, daily	(5) Custom., daily
Equipment (T1 or T2)	0.250***	0.962***	0.047*	0.072**	0.059*
	(0.021)	(0.064)	(0.022)	(0.027)	(0.029)
	[0.059]	[0.034]	[0.059]	[0.034]	[0.074]
Vendors 0-400m	0.005	0.027***	0.007**	0.005	0.004
	(0.003)	(0.007)	(0.002)	(0.003)	(0.003)
	[0.032]	[0.097]	[0.032]	[0.097]	[0.128]
Vendors 400-800m	0.005	0.033***	0.008*	0.006	0.006
	(0.003)	(0.008)	(0.003)	(0.004)	(0.005)
	[0.059]	[0.109]	[0.059]	[0.109]	[0.122]
T1 or T2 vendors 0-400m	-0.007	-0.040***	-0.012**	-0.011*	-0.008
	(0.004)	(0.012)	(0.004)	(0.005)	(0.005)
	[0.027]	[0.059]	[0.027]	[0.059]	[0.080]
T1 or T2 vendors 400-800m	-0.008	-0.051***	-0.012*	-0.009	-0.011
	(0.005)	(0.013)	(0.006)	(0.007)	(0.008)
	[0.059]	[0.122]	[0.059]	[0.122]	[0.122]
Control mean:	7.17	0.88	9.67	11.34	7.46
Clusters:	108	108	108	108	108
Observations:	3587	3587	3102	3105	2892

Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcome variable definitions and estimation details can be found in Tables 7 and 8. We aggregate treatment groups for convenience. "Vendors 0-400m" is calculated as the number of vendors within a 0 to 400m radius (not mean-centered); "Vendors 400-800m" is defined similarly. All regressions include strata fixed effects and the pre-treatment average sanitary equipment observed at the kiosk. To increase precision, we also include fixed effects for the survey period, interviewer, number of employees, and years of experience. Standard errors at the cluster level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A43: Treatment Effects on Quality and Business Outcomes with Weighted Spillover Effects

	(1) Overall quality (0-20)	(2) Large equip. (0-4)	(3) Profits, monthly	(4) Sales, monthly	(5) Custom., monthly
Equipment (T1 or T2)	0.253***	0.989***	0.055*	0.078**	0.067*
	(0.020)	(0.062)	(0.023)	(0.026)	(0.030)
	[0.080]	[0.031]	[0.080]	[0.031]	[0.080]
Total weighted vendors	0.001	0.004	0.001	0.001	-0.000
	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)
	[0.308]	[0.274]	[0.308]	[0.274]	[0.604]
T1 or T2 weighted vendors	0.000	-0.001	-0.003	-0.005	-0.002
	(0.002)	(0.005)	(0.003)	(0.003)	(0.004)
	[0.274]	[0.117]	[0.274]	[0.117]	[0.532]
Control mean:	7.17	0.88	9.67	11.34	7.46
Clusters:	108	108	108	108	108
Observations:	3587	3587	3102	3105	2892

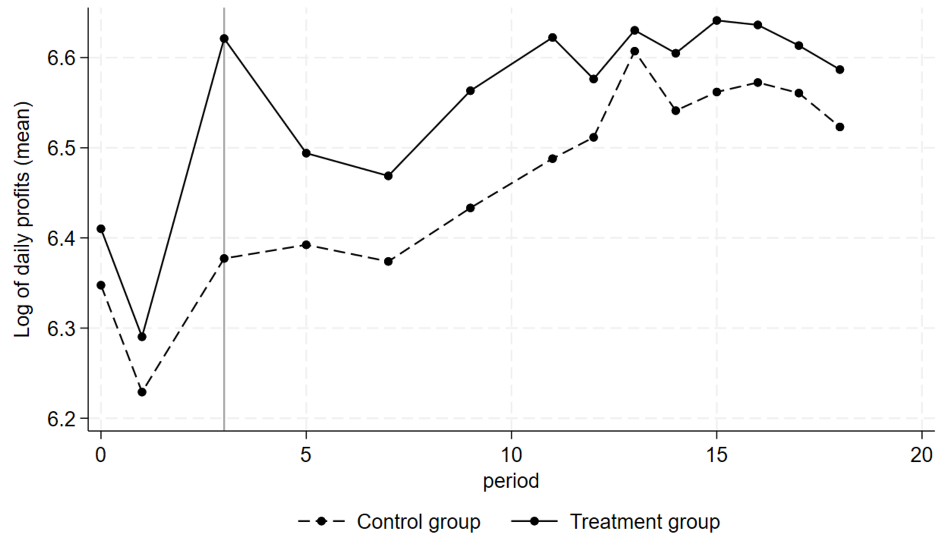
Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcome variable definitions and estimation details can be found in Tables 7 and 8. We aggregate treatment groups for convenience. "Vendors 0-400m" is calculated as the weighted number of vendors within a 0 to 400m radius; "Vendors 400-800m" is defined similarly. Weights are equal to $1/(1+distance)$. All regressions include strata fixed effects and the pre-treatment average sanitary equipment observed at the kiosk. To increase precision, we also include fixed effects for the survey period, interviewer, number of employees, and years of experience. Standard errors at the cluster level. $^*p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

Table A44: Treatment Effects on Quality and Business Outcomes with Square Weighted Spillover Effects

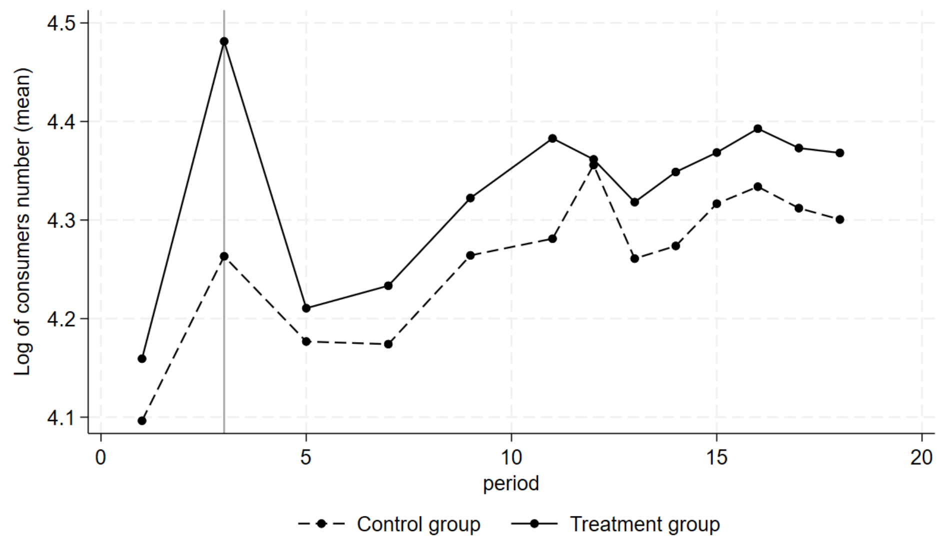
	(1) Overall quality (0-20)	(2) Large equip. (0-4)	(3) Profits, monthly	(4) Sales, monthly	(5) Custom., monthly
Equipment (T1 or T2)	0.253***	0.989***	0.055*	0.079**	0.067*
	(0.020)	(0.062)	(0.023)	(0.026)	(0.030)
	[0.075]	[0.028]	[0.075]	[0.028]	[0.077]
Total sq-weighted vendors	0.001	0.004	0.001	0.001	-0.000
	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)
	[0.262]	[0.241]	[0.262]	[0.241]	[0.538]
T1 or T2 sq-weighted vendors	0.001	-0.002	-0.005	-0.007	-0.003
	(0.002)	(0.006)	(0.003)	(0.004)	(0.005)
	[0.241]	[0.099]	[0.241]	[0.099]	[0.538]
Control mean:	7.17	0.88	9.67	11.34	7.46
Clusters:	108	108	108	108	108
Observations:	3587	3587	3102	3105	2892

Notes: Data from monitoring surveys (random audits) measured during the entire study period (including periods 3 and 4 when the equipment was being delivered) up until the second endline survey in February 2023. Outcome variable definitions and estimation details can be found in Tables 7 and 8. We aggregate treatment groups for convenience. "Vendors 0-400m" is calculated as the squre-weighted number of vendors within a 0 to 400m radius; "Vendors 400-800m" is defined similarly. Weights are equal to $(1/(1+distance))^2$. All regressions include strata fixed effects and the pre-treatment average sanitary equipment observed at the kiosk. To increase precision, we also include fixed effects for the survey period, interviewer, number of employees, and years of experience. Standard errors at the cluster level. $^*p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$. Sharpened q-values (Benjamini et al., 2006) are in brackets.

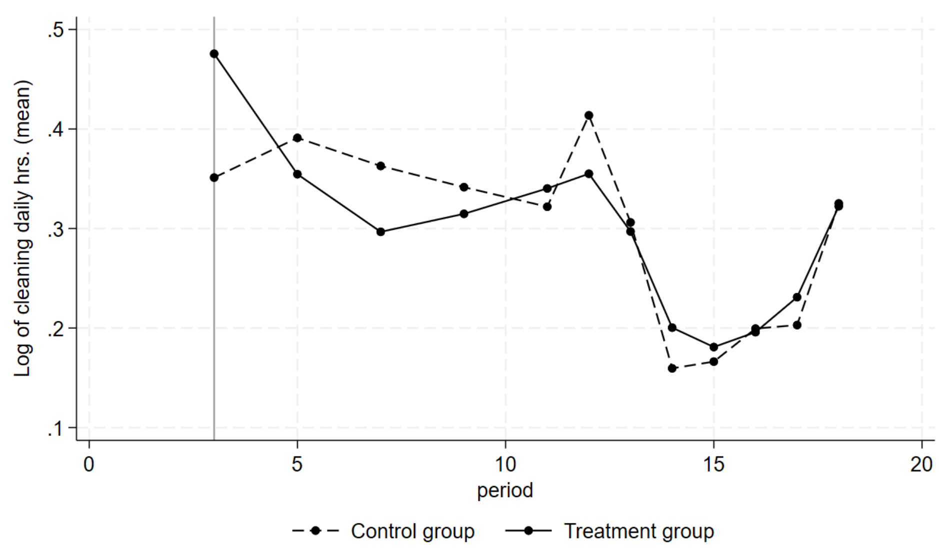
Figure A17: Changes in Profits, Customer Numbers, and Cleaning Over Time



a) Vendor Profits



b) Customer Numbers



c) Hours Spent Cleaning

Notes: Data from monitoring and endline surveys. Figures are produced using average customers number, average profits and average time dedicated to cleaning across the study period for treatment and control groups. The vertical line indicates the start of the study period. The treatment period ended in period 15 (3 baseline + 12 treatment weeks).

J Conceptual Framework: Extensions and Discussion

In the main text, we consider a simplified model of moral hazard in which detection is captured by a single parameter λ: a vendor who merely pretends to be high quality loses automatically the λ fraction of consumers who detect shirking, serving only the remaining fraction $(1-\lambda)$. That simpler assumption yields the separating condition $p\lambda + c_0(1-\lambda) > c_s$. Here, we provide a more general framework in which *all* consumers receive a *noisy* signal about quality, rather than having a fixed share λ who see through low effort, but with different likelihoods of "looking high-quality" under H vs. L. This approach is more in line with Bayesian updating and signal-detection models commonly used to capture imperfect information in markets.

J.1 Setup

Let H and L index whether the vendor produces high- or low-quality food. High-quality production requires the hidden per-unit cost $c_s > c_0$, whereas low quality is produced at baseline cost c_0. However, instead of detection being captured by a fraction of consumers who automatically see through low quality, we now let each consumer observe a random signal s, drawn from

$$s \sim \begin{cases} F(\cdot \mid H), & \text{if the vendor is high quality,} \\ F(\cdot \mid L), & \text{if the vendor is low quality.} \end{cases}$$

Each consumer updates beliefs about the vendor's quality using Bayes' rule. Concretely, let $\ell(s) = \frac{f(s|\mathrm{H})}{f(s|\mathrm{L})}$ be the likelihood ratio at signal s. Suppose a consumer is willing to buy at the high-quality price p_H if and only if $\ell(s)$ exceeds a certain threshold ℓ^*. Define

$$\Phi_{\mathrm{H}} \;=\; \Pr(\ell(s) > \ell^* \mid \mathrm{H}), \quad \Phi_{\mathrm{L}} \;=\; \Pr(\ell(s) > \ell^* \mid \mathrm{L}).$$

Hence, a truthful H vendor convinces a fraction Φ_{H} of consumers to purchase at p. Conversely, a shirking L vendor fools only Φ_{L} of consumers. This contrasts with the main text's simpler assumption that a vendor who merely pretends to be high quality sells to the remaining $(1-\lambda)$ share of buyers. If we impose $\Phi_{\mathrm{H}} = 1$ and $\Phi_{\mathrm{L}} = 1-\lambda$, we immediately recover the simpler fraction-detection model. In reality, of course, partial detection or false alarms across *all* consumers is often more realistic than the "bang-bang" λ fraction approach.

J.2 Equilibrium Incentives

Treated vendors already possess the visible equipment giving access to the higher demand curve $D_H(p)$. Their per-period profits are

$$\pi_H^T(p) = \big[p - c_s\big]\,\Phi_H D_H(p) \quad \text{versus} \quad \pi_L^T(p) = \big[p - c_0\big]\,\Phi_L D_H(p).$$

mirroring the main text's approach but replacing λ with the more general $\Phi_{\mathrm{H}}, \Phi_{\mathrm{L}}$. High-quality production is incentive-compatible when

$$\pi_H^T(p) > \pi_L^T(p) \iff \big[p - c_s\big]\Phi_H > \big[p - c_0\big]\Phi_L.$$

Equivalently,

$$p\,(\Phi_H - \Phi_L) > (c_s - c_0)\,\Phi_H,$$

so the *signalling advantage* $\Phi_H - \Phi_L$ that genuine high-quality producers enjoy must be large enough relative to the incremental cost $c_s - c_0$. Equivalently, the signal distributions $F(\cdot \mid \mathrm{H})$ and $F(\cdot \mid \mathrm{L})$ must differ enough that truly high-quality vendors consistently appear more convincing to most consumers.

J.3 Type 1 and Type 2 Errors: Discussion

The reason a stochastic signal framework may be preferable is that real-world "detection rates" often conflate multiple types of errors and do not map neatly into a single fraction λ. In standard signal-detection theory, there are two key error probabilities:

1. **Type 1 Error** (α): The probability of *incorrectly* flagging a truly high-quality vendor as low quality (i.e. a "false positive"). In our notation,

$$\alpha \;=\; \Pr\bigl[\ell(s) \le \ell^* \bigm| \mathrm{H}\bigr],$$

   so that $\Phi_{\mathrm{H}} = 1-\alpha$ is the fraction of signals that still look "convincingly high quality" to consumers when the vendor is genuinely HQ.

2. **Type 2 Error** (β): The probability of *failing* to detect a low-quality vendor who is pretending to be high quality (i.e. a "false negative"). We have

$$\beta \;=\; \Pr\bigl[\ell(s) > \ell^* \bigm| \mathrm{L}\bigr],$$

   implying that $\Phi_{\mathrm{L}} = \beta$ is the fraction of consumers who are fooled into paying for high quality when the vendor is actually low quality.

This distinction matters because the presence of Type 1 errors reduces the expected profit of *true H* vendors, as some fraction of signals incorrectly classify them as L (resulting in lost sales). At the same time, Type 2 errors allow some fraudulent L vendors to go undetected, raising their profits and undermining overall market quality. Our condition

$$p\,(\Phi_H - \Phi_L) > (c_s - c_0)\,\Phi_H,$$

can thus be written in terms of α and β as:

$$\bigl[p - c_s\bigr]\,(1-\alpha) \;>\; \bigl[p - c_0\bigr]\,\beta.$$

If α is large, even honest vendors appear suspicious to many consumers, shrinking their profit advantage. If β is large, too many low-quality fakers manage to deceive buyers. In either case, sustaining high-quality production in equilibrium becomes more difficult.

J.4 Interpreting "41% of Consumers Detect Unsafe Food"

Our survey reports that "41% of consumers detect unsafe food." This statistic is, at best, an incomplete measure of detection efficacy because it does *not* distinguish between Type 1 and Type 2 errors. For instance, it may reflect $\Pr[\text{detect L} \mid \text{L}] = 1 - \beta = 0.41$, which is the probability that a consumer *correctly* identifies a low-quality vendor. This would imply a Type 2 error rate of 59%. Such an interpretation aligns with the simplifying assumption in the main text that $\alpha = 0$, leading to:

$$\Phi_{\text{H}} = 1, \quad \Phi_{\text{L}} = 1 - \lambda,$$

where λ represents the fraction of consumers who perfectly detect low effort. While this simplification collapses detection into a single parameter, it overlooks the possibility that consumers may frequently misclassify *high-quality* vendors as low quality ($\alpha > 0$). Without knowing the prevalence of Type 1 errors, we cannot accurately assess detection capabilities. Therefore, the survey's "41% detection" claim cannot be directly interpreted as $\lambda = 0.41$ in the simple model, nor does it allow us to recover Φ_{H} and Φ_{L} in the more general setting without additional data on *both* Type 1 and Type 2 errors.

J.5 Price Coordination and Fixed-Price Profits

In our endline survey, 67% of vendors reported that prices are set collectively and 86% said deviating from the norm is disapproved of by peers, implying that every kiosk charges the same menu price p. When a treated vendor upgrades, consumers perceive higher quality, so demand jumps from $D_L(p)$ to $D_H(p) > D_L(p)$ while the price itself cannot change. The resulting profit difference is

$$\Delta\pi = \underbrace{[p - c_s] D_H(p)}_{\text{profit after upgrading}} - \underbrace{[p - c_0] D_L(p)}_{\text{baseline profit}} = \Big[p - (c_s - c_0)\Big]\big[D_H(p) - D_L(p)\big].$$

Thus profits rise only because more units are sold at the same price; the gain is tempered by the higher per-unit cost $c_s - c_0$ once subsidies end.

Because kiosk space and labour are limited, the quantity gap $D_H(p) - D_L(p)$ cannot grow indefinitely. With capacity constraints capping extra sales and a fixed price preventing mark-ups, the net benefit of maintaining safe practices shrinks rapidly after subsidies expire—exactly matching the post-intervention reversals documented in Table 7.