Transforming Brick Manufacturing in Bangladesh to Promote Clean Air and Better Health

Pre-Analysis Plan

March 16, 2023

This document outlines the analysis plan for a randomized controlled trial of technical training and incentives for brick kiln owners, part of "Transforming Brick Manufacturing in Bangladesh to Promote Clean Air and Better Health." The document provides a pre-specified methodology and plan for analyzing the results of the experiment.

1 Study Overview

Across South Asia, the brick manufacturing industry is dominated by inefficient, coal-burning kilns. Brick kilns are one of the largest emitters in the region. In Bangladesh, kilns contribute 17% of the country's annual CO2 emissions and 11% of PM2.5. The pollution released by these kilns worsens local air quality, health and agricultural productivity, and global climate. Moreover, brick production is known for poor working conditions, including indicators of labor trafficking such as debt bondage, excessive work requirements, and hazardous working conditions. Reducing emissions and improving work conditions could generate large social benefits. A properly constructed and operated zigzag kiln can reduce black carbon by 41%, CO2 by 21%, and PM2.5 by 80% – and strikingly, also increase kiln profitability. Yet, the vast majority are incorrectly constructed and operated. Our preliminary work found that lack of knowledge regarding proper construction and operation and inattention to worker incentives undermine kiln operation. We propose a randomized intervention among kiln owners in Bangladesh that relaxes these barriers to improve kiln performance, reduce air pollution, lower greenhouse gas emissions, and reduce the prevalence of trafficking indicators.

2 Experimental Design

A properly constructed and operated ZZK can reduce black carbon by 41%, CO2 by 21%, and PM2.5 by 80%. However, most conversions to ZZKs in Bangladesh have been poorly implemented. Consequently, the new ZZK appear to be just as polluting as the older style they replaced. Our study of kiln owners suggests a puzzle: a correctly built and operated ZZK can increase kiln profits, yet these gains are not realized by owners with poorly constructed and operated kilns. Our preliminary work suggests two primary barriers to effective implementation: 1) lack of knowledge of the specific interventions and their true economic return and 2) inattention to the incentives of workers whose cooperation is crucial to running an efficient kiln.

We will address these barriers through a randomized intervention that provides:

1. Extensive information, training, and technical support on low-cost improvements
2. Information and encouragement to kiln owners to adopt strategies that incentivize workers to adopt the new practices

2.1 Technical Only Arm

Kilns assigned to the technical arm will receive information, training, and technical support to adopt a suite of technical and operational improvements including improved firing practices, improved brick setting, increased insulation, and good bookkeeping. The trainings will highlight the financial benefits of these improvements and include live participation from owners who adopted them during our pilot, to directly address owners' uncertainty of economic returns. Table 1 describes the five components of the technical intervention. The first two components (double/triple zigzag stacking and single fireman continuous feeding) are the two most important of the five elements for achieving reductions in emissions, particulate matter, and coal use. Kiln owners that adopt these two practices after their first round of bricks are fired will be given a banner that promotes their participation in the project.

The trainings will be delivered in the form of initial orientation sessions for owners and their managers, as well as a subsequent session for the firing and loading *sardar* (labor supervisors). These initial sessions will be followed by on-site trainings of the sardar and workers on brick loading and firing. Throughout the brick firing season, our technical team will provide technical support to help owners and their workers implement the new practices.

2.2 Technical + Incentive Arm

Kilns assigned to the incentive group will receive everything delivered to the technical arm, in terms of technical training and support, plus additional information and encouragement targeted toward owners to address the workers' misaligned incentives. This will include a preliminary group meeting with all owners assigned to the incentive arm, in which our team will explain the workers' incentives and how they are misaligned with the owners' incentives, the importance of workers properly adhering to the new technical practices and giving them enough time, training, and positive reinforcement to adopt the new practices, and suggestions of several ways to incentivize workers to adhere to these new practices. These suggestions are informed by the economics and social psychology literature, as well as the experience of Indian kiln owners and kiln owners from our pilot study in Jashore district, and include a mix of financial (e.g., bonuses, higher wages, return bonuses) and non-financial incentives (e.g., better working conditions, such as meals, housing, clothing).

A handout describing the importance of motivating workers to adopt the new technical practices will be left with the owners at this initial meeting. Our team will conduct two follow-up "nudge" visits with all kilns in the incentive arm. At the first follow-up nudge visit, they will give owners a poster that presents a few simplified, key messages around the importance of incentives as a reminder.

Table 1: Technical Interventions to Improve Kiln Performance

	Intervention	Rationale
1	Changing the pattern of brick stacking into double/triple zigzag	Improves airflow which improves the completion of combustion, reducing black carbon and small particulates
2	Single firemen continuous feeding	Reduces the amount of coal per feed, and so improves the completeness of combustion
3	Closing wicket gates with cavity wall	Reduces heat loss, which reduces the total amount of coal that needs to be combusted per 1,000 bricks produced
4	Thicker ash layer	Reduces heat loss, which reduces the total amount of coal that needs to be combusted per 1,000 bricks produced
5	Use of sawdust/biomass in front chambers	Helps achieve more complete combustion of coal and reduces emissions of black smoke/carbon

2.3 Control Arm

Kilns assigned to the control group will receive no intervention, information, or encouragement, but will be visited by our team for all data collection (baseline, efficiency assessment, endline).

2.4 Sampling and Randomization

We are conducting this experiment in Khulna Division of Bangladesh. To select kilns for inclusion in the trial, we began by contacting the Brick Manufacturing Owners Associations within each district of Khulna Division. We had initial discussions with each district association to find out how many coal-fired zigzag kilns were operating within their district, which resulted in a total of 410 kilns. We obtained lists of all zigzag kilns using coal (instead of firewood or other biomass), along with contact information, for each district and contacted them to explain our study, including the randomized design. We conducted a baseline survey of owners in 355 kilns[1] After describing the study, we obtained initial interest in participation and enrolled kilns in the study with the ultimate aim of 300 kilns (see Section 3.5 for power calculations below).

After collecting baseline data from 355 kilns(see Section 2.5), we restricted the sample to owners with only 1 kiln (e.g., excluding any additional kilns owned by someone in the sample) and ensured that firing and loading sardar worked for at most 2 kilns in the sample. These restrictions eliminated 27 kilns, leaving us with a sample of 328 kilns.

---

[1] Of the 410 kilns initially provided, we learned that 55 were not actually zigzag kilns, not being operated during the current firing season, or the owners were planning to use only firewood and therefore excluded from our study.

We randomized kilns to one of the three arms stratifying by district and baseline quality of bricks (from the previous season), where kilns were classified as above/below the median % of class 1 bricks. We conducted a re-randomization process in which we constructed 1,000 different randomization allocations and checked for balance on a set of key characteristics (Morgan and Rubin, 2012). We identified the allocations that achieved acceptable balance as those in which none of the hypothesis tests were statistically different from 0 at the 5% level. Among the acceptable allocations, we selected the one with the largest sum of all p-values for our final allocation.[2]

The final allocation of kilns includes 108 kilns assigned to the control arm, 108 assigned to the technical only arm, and 112 assigned to the technical + incentive arm. The sample was balanced (see Table 2) across all 3 treatment arms on 15 variables of interest: kiln owner experience, kiln owner education, joint ownership (e.g., a joint proprietorship), knowledge of the Jashore intervention, interaction with a Jashore pilot kiln, year of conversion to zigzag, whether the kiln is adjacent to water, whether the kiln is located on highland, the total lakh bricks produced in the last season, total circuits completed in the last season, % of class 1 bricks produced in the last season, cost of production per 1000 bricks, weight of fired bricks, total number of workers, and whether the kiln has a shared sardar (either loading or firing).

Table 2: Balance Tests for Original Sample of 328 Kilns

Balance Variable	Incentive Mean	Incentive Std. Dev.	Technical Mean	Technical Std. Dev.	Control Mean	Control Std. Dev.	I - C (p-val)	T - C (p-val)	I - T (p-val)
Owner Experience (Years)	14.134	8.225	15.528	10.104	14.343	9.665	0.820	0.397	0.247
Jashore Intervention Knowledge	0.294	0.458	0.321	0.470	0.284	0.454	0.985	0.567	0.553
Jashore Owner Interaction	0.480	0.510	0.500	0.510	0.609	0.499	0.686	0.794	0.861
Zigzag Year	2014.670	3.514	2014.176	4.277	2014.389	3.716	0.617	0.671	0.360
Water Adjacent	0.616	0.489	0.602	0.492	0.620	0.488	0.955	0.789	0.836
Bricks Fired (Lakhs)	7.978	1.039	8.037	1.212	8.010	1.231	0.803	0.889	0.685
Circuits Completed	5.945	1.630	5.762	1.497	5.869	1.800	0.758	0.607	0.368
Class 1 Production Share (%)	64.562	11.447	65.926	10.505	64.935	10.756	0.852	0.327	0.259
Production Cost Estimate BDT (per 1K Bricks)	8754.375	1155.422	8529.815	1192.925	8666.204	1032.754	0.506	0.358	0.139
Fired Brick Weight (kg)	3.369	0.250	3.408	0.222	3.408	0.230	0.193	0.983	0.205
Total Workers	108.607	27.821	107.750	31.538	109.343	34.215	0.909	0.735	0.806
Higher SEC Plus	0.598	0.492	0.611	0.490	0.602	0.492	0.968	0.875	0.905
Highland	0.723	0.449	0.722	0.450	0.722	0.450	0.990	0.967	0.956
Joint Ownership	0.312	0.466	0.343	0.477	0.370	0.485	0.336	0.679	0.583

---

[2] Due to high coal prices in 2022, many of our originally included kilns decided to switch to exclusive firewood use after the baseline data was collected, which excludes them from participating in the technical intervention. To supplement our initial sample, we later enrolled an additional 30 kilns from Jashore district, assigning 10 to each arm. Among these additional 30 kilns, we conducted the same re-randomization procedure and tested for balance both within the 30 new kilns and across all 124 kilns from Jashore. We selected a randomization allocation that was acceptable in both cases and achieved the highest sum of p-values across all tests when we considered the full 124 sample of kilns. We will present all analyses for the original sample of 328 kilns and the expanded sample of 358 kilns. See Table 3 for the balance tests on the combined sample of 358 kilns.

Table 3: Balance Tests for Combined Sample of 358 Kilns

Balance Variable	Incentive Mean	Incentive Std. Dev.	Technical Mean	Technical Std. Dev.	Control Mean	Control Std. Dev.	I - C (p-val)	T - C (p-val)	I - T (p-val)
Owner Experience (Years)	14.393	8.100	15.681	9.946	14.846	9.823	0.665	0.534	0.255
Jashore Intervention Knowledge	0.368	0.485	0.370	0.485	0.356	0.481	0.963	0.840	0.875
Jashore Owner Interaction	0.457	0.505	0.529	0.507	0.531	0.507	0.854	0.683	0.559
Zigzag Year	2014.598	3.430	2014.227	4.132	2014.376	3.617	0.684	0.765	0.486
Water Adjacent	0.607	0.491	0.605	0.491	0.624	0.486	0.773	0.829	0.940
Bricks Fired (Lakhs)	7.959	1.012	7.975	1.227	8.010	1.198	0.696	0.810	0.894
Circuits Completed	5.982	1.607	5.835	1.509	5.913	1.790	0.761	0.672	0.425
Class 1 Production Share (%)	65.008	11.084	66.176	10.120	65.239	10.462	0.929	0.370	0.343
Production Cost Estimate BDT (per 1K Bricks)	8782.705	1114.078	8573.277	1157.054	8695.299	997.648	0.485	0.357	0.131
Fired Brick Weight (kg)	3.370	0.241	3.405	0.214	3.403	0.224	0.236	0.942	0.211
Total Workers	107.836	27.830	108.647	31.138	108.487	33.851	0.918	0.954	0.866
Higher SEC Plus	0.598	0.492	0.597	0.493	0.581	0.495	0.707	0.783	0.921
Highland	0.738	0.442	0.723	0.450	0.735	0.443	0.966	0.747	0.709
Joint Ownership	0.311	0.465	0.328	0.471	0.376	0.486	0.269	0.441	0.740

2.5 Data Collection

We developed four sets of quantitative data collection tools:

1. kiln owner questionnaire
2. adoption checklist
3. kiln performance assessment
4. worker survey

The kiln owner questionnaire was completed at baseline in October 2022 before randomization, before the start of the brick firing season in November, and before the delivery of the intervention. The kiln owner questionnaire collects information on kiln owner demographics, the GPS location of the kiln, retrospective information on the previous brick firing season (production, costs, revenue, labor, use of incentives, working conditions), and information about kiln construction and operation. The same questionnaire will be fielded at endline in June-July 2023, after the kilns have completed firing.

The adoption checklist will be a short, technical checklist to assess whether or not kilns have adopted the five technical intervention practices that are listed in Table 1. All kilns will be visited after they have fired the first brick round of bricks, approximately in late January.

The kiln performance assessment tool will be fielded in April 2023, after kilns have completed several rounds of brick firing and the process has reached an “equilibrium” in terms of energy use (typically more coal is used in early rounds of firing). The efficiency assessment will take approximately 24 hours per kiln and will collect data for the primary outcomes for the trial. The assessment will involve counts of brick production by brick class, making flue gas measurements, making coal measurements and sampling coal, and tracking the pace of fire travel. The adoption checklist will also be conducted during the 24-hour efficiency assessment to provide another data point on adoption, however, banners will not be given out to owners who have adopted at this stage.

We will also conduct an endline survey of workers, which will allow us to identify whether working conditions, incentives, or benefits will differ across the three arms and also differ from what owners report in the kiln owner survey. Six workers across four categories of employment (moulding, loading, firing, unloading) will be surveyed, along with two sardars (recruiters and team managers) in two different work categories. The worker survey will collect information about their work conditions, age range of workers at the kiln, socioeconomic status, migration status, compensation and incentives received, occupational hazards, and physical and mental health indicators. Survey's conducted among sardars are modified to collect relevant information about team management practices and owner-sardar interactions. The worker survey will be conducted in May-June 2023.

3 Empirical Strategy

In order to estimate the treatment effects of the intervention we will regress outcomes on dummy variables indicating treatment status.

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 I_i + \gamma_{s(i)} + \epsilon_i \tag{1}$$

where Y_i is an outcome of interest for kiln i, T_i is a binary indicator for assignment to the technical only arm, I_i is a binary indicator for assignment to the incentive arm and $\gamma_{s(i)}$ is an indicator for the stratum for observation i. The coefficients on each treatment indicator, β_1 and β_2, respectively, capture the "intention-to-treat" (ITT) effect of assignment to the treatment arms on each of the outcomes relative to the control arm.

We will also estimate a version of Equation 1 in which we bundle treatment into a single treatment indicator that captures the ITT effect of assignment to either treatment arm (Equation 2). Heteroscedasticity-robust standard errors will be calculated for all specifications.

$$Y_i = \delta_0 + \delta_1 G_i + \gamma_{s(i)} + \epsilon_i \tag{2}$$

Because we do not expect all 200 kilns assigned to the treatment arms will adopt the technical intervention (where adoption is defined as taking up both of the recommended brick stacking and firing practices) and we do not expect control firms to implement any of the interventions (i.e. one-sided compliance) we will also estimate a "treatment-on-the-treated" (TOT) specification.

This specification allows us to quantify the impact of the technical intervention among the kilns that *actually* took up the recommended practices. Because non-adoption and noncompliance are not random but likely the result of systematic differences between kiln owners that are likely correlated with the outcomes, we will use use random assignment as an instrument for adoption in an instrumental variables analysis that estimates the local average treatment effect among the kilns that took up the intervention (e.g., the compliers).

To estimate the TOT, we used the following two-stage least squares (2SLS) approach:

$$\text{Adopt}_i = \theta_0 + \theta_1 G_i + \gamma_{s(i)} + \epsilon_i \tag{3}$$

$$Y_i = \gamma_0 + \gamma_1 \text{Adopt}_i + \gamma_{s(i)} + u_i \tag{4}$$

Equation 3 is the first stage in which adoption (or take-up) of the two most critical intervention components (double or triple zigzag brick setting and single fireman continuous coal feeding) is predicted with the randomly assigned treatment (using a bundled treatment indicator, G_i). Equation 4 is the outcome equation and is estimated by 2SLS. Heteroscedasticity-robust standard errors will be calculated for all specifications, clustering at the kiln level for outcomes measured at the individual worker level.

We will also carry out complementary analyses for the two specific treatment arms. First, we can compute arm-specific LATE estimates (i.e. estimating LATE for control versus T and for control versus I). Second, we estimate a single 2SLS using (T, I) as instruments. This procedure yields a weighted average of pairwise treatment effects that have a meaningful economic interpretation under certain conditions.[3]

We will present all analyses for the original sample of 328 kilns and the expanded sample of 358 kilns.

3.1 Primary Outcomes & Hypotheses

Primary outcomes will include carbon monoxide/carbon dioxide ratio (a measure of combustion efficiency measured by placing a sensor into the flu gas), specific energy consumption (the energy used in MJ for firing 1 kg of brick) and proportion of class 1 bricks produced. We will track coal usage and sample coal to test for carbon content, and then apply the IPCC's methodology to estimate tons of CO2 abated due to the intervention by converting specific energy consumption to CO2.

3.1.1 CO/CO2

The CO/CO2 ratio is a measure of complete combustion – the more complete the combustion, the more carbon gets converted into CO2 and the lower the ratio. It is measured by collecting data with a flue gas analyzer. A CO/CO2 ratio ≤ 0.025 is indicative of more efficiency zigzag kiln operation (using coal as the main fuel). The CO/CO2 ratio is a cross sectional assessment of the combustion efficiency at the time of measurement, rather than an averaged performance for the whole season.

3.1.2 Specific Energy Consumption (SEC)

Specific energy consumption (SEC) is a scientific measure of energy performance that takes into account variation in gross calorific values of coal and fired brick weight. A lower SEC indicates higher efficiency. It is calculated by multiplying the tons of coal consumed by the gross calorific value (in MJ/kg) and dividing by the total production of bricks (in kg), thus the resulting units of specific energy consumption are in MJ/kg-fired brick.

3.1.3 Class-1 bricks

Brick production and quality are quantified by summing the total number of bricks fired during completed batches and multiplying by the proportion of bricks that fall into the following classes: class-1, class-1.5, class-2, class-3, and breakages. Class-1 represents the highest quality of bricks

---

[3] See e.g. Mogstad et al. (2020); Imbens and Angrist (1994) on estimation of treatment effects with non-binary or vector valued instruments.

and command the highest price. To produce Class-1 bricks, kilns need to operate with consistent temperature throughout the firing zone. Thus, the share of class-1 bricks is an indicator both of higher profit and energy efficiency which would generate fewer emissions.

3.2 Secondary Outcomes & Hypotheses

Secondary outcomes include additional measures of energy efficiency, brick production and profitability, as well as measures of working conditions and whether any type of worker incentives were provided. A secondary energy efficiency outcome we will assess is the specific fuel consumption (SFC), which is a common metric used by the brick kiln owners to estimate the efficiency of fuel use and calculated as the total quantity of fuel consumed divided by the total quantity of bricks produced. We will also follow standard guidelines to convert the primary outcome specific energy consumption into its CO2 equivalent by applying the Intergovernmental Panel on Climate Change (IPCC) carbon emissions factor for fuel, and a carbon to CO2 conversion factor.

Although the proportion of class-1 bricks is important for both efficiency and profitability, the intervention may also induce a shift in the entire distribution of bricks from lower classes to higher class. Thus, as a secondary brick quality outcome we will include the proportion of bricks in all other classes.

Because the technical interventions increase the energy efficiency of kilns and coal is the largest input cost for owners, as a secondary outcome we will calculate the total amount of money spent on coal, as well as the average coal spending per brick.

While a reduction in coal use and increase in class-1 bricks imply increased profit (through both lower costs and higher revenue), measuring profit directly is complicated because the decision of when to sell bricks is an endogenous choice by owners – who may keep stocks of bricks to wait for better prices. Therefore, measures of revenue or profit are likely to be biased. To address the issue, we will calculate the value of production by multiplying reported brick prices (using the median price per class reported by all kiln owners) by production. We will calculate two measures: overall value of production and the average value of production per brick.

3.2.1 Working Conditions

Outcomes related to working conditions are designed to measure incremental improvements in employment practices that contribute to trafficking. A worker survey conducted among six workers and two sardars across four categories of work (moulding, loading, firing, unloading) at each site documents the presence of 18 strong and 20 medium indicators of labor trafficking across 7 categories of worker exploitation (deceptive recruitment, employment practices and penalties, personal life and property, degrading conditions, freedom of movement, debt and dependency, and violence or threats of violence) (see Okech et al., 2020 for detailed description of each indicator).

Four outcomes measure trafficking-related working conditions: 1) a binary outcome capturing whether or indicators at a given kiln site are sufficient to constitute labor trafficking (Okech et al. 2020), 2) a binary indicator for the presence of the worst forms of child labor, defined as children working in jobs that meet the labor trafficking criteria, 3) a continuous outcome equivalent to the number of individual indicators present at a given brick kiln work site, weighting medium indicators at

2/3 relative to strong indicators, and 4) a continuous measure of the number of individual trafficking indicators directly targeted by the suggested worker incentive structures. Of the indicators specified in Okech (2020), three are most closely related to the incentives. These include *DC1, Made to be available day and night without adequate compensation outside of the scope of the contract; DC2, Made to complete hazardous and/or arduous services without proper protective gear; and DC4, Made to live in degrading conditions (e.g. housing or shelter is unclean, provides no privacy, or is otherwise insufficient in a way that harms health)*.

3.3 Heterogeneity

We will explore heterogeneity in the primary outcomes across dimensions such as kiln owner age, years of experience in the brick industry, whether the kiln owner is involved in other businesses, baseline brick production, and kiln location.

3.4 Multiple Hypothesis Testing

We will apply corrections for multiple hypothesis testing to the following families of hypotheses below. For each family, we will implement Holm corrections as well as the bootstrap procedure outlined in List et al. (2016) and the procedures in Benjamini and Hochberg (1995), Anderson (2008), Romano and Wolf (2005) and Romano and Wolf (2016).

1. Heterogeneity in Primary Outcomes: We will compute heterogeneity in the ATE for each of the primary outcomes by the covariates listed above. For each primary outcome and for each covariate, we will implement a multiple test correction for testing the null hypotheses jointly (i.e. that the ATE is zero for each sub-group defined by different values of the covariate).

2. Working Conditions: We will evaluate the ATEs for the four trafficking related outcomes jointly using the multiple test corrections above.

3.5 Power calculations

Based on our pilot results, we have estimated effect sizes for the “intention-to-treat” (ITT) effect of each experimental arm, as well as a “treatment-on-the-treated” (TOT) effect that accounts for imperfect compliance with the intervention (both from kilns assigned to the treatment arm that did not take-up the intervention practices and from control kilns that sought to learn the intervention practices) by using random assignment to both arms as an instrument for adoption. These results for each of the three outcomes are summarized in Table 4 below. We first calculate the minimum detectable effect size (MDES) assuming both arms have equal effect sizes, a significance level of 0.05 and power of 0.9. Then, because there is suggestive evidence from our pilot that the incentive arm encouraged better adherence to the improved operating practices and resulted in better outcomes, we also calculate our statistical power for detecting differences between the incentive and technical arms.

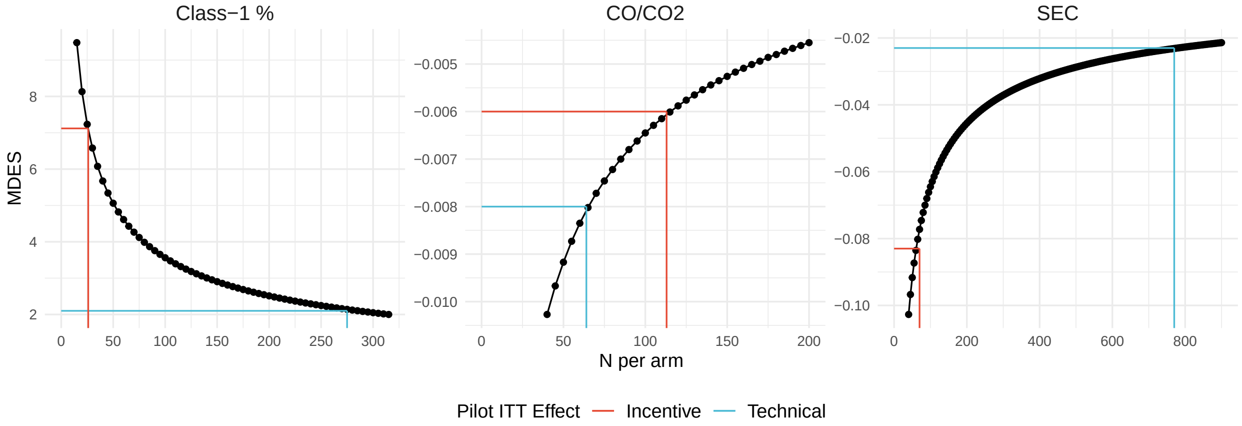
Figure 1 presents the minimum detectable effect sizes against the sample size per treatment arm for the percent of class-1 bricks produced, CO/CO2 ratio, and specific energy consumption. The estimated ITT effects for each arm from the pilot study are indicated in red (incentive arm) and blue (technical arm). These scenarios indicate that with a sample size of 100 kilns per experimental arm

Table 4: Estimated Effect Sizes from Pilot Study

Outcome	**Control Group Mean**	**Technical ITT**	**Incentive ITT**	**TOT**
Class-1 (%)	66	2.1	7.12	9.22
CO/CO2 (ratio)	0.04	-0.008	-0.006	-0.014
SEC (MJ/kg-fired brick)	1.28	-0.023	-0.083	-0.107

(300 total kilns), we are powered for all three outcomes with 90% power in most cases. For class-1 bricks the incentive arm performed much better, producing 7.12 percentage points more class-1 bricks than the control group and we would be powered to detect an effect size of this magnitude with only 25 kilns per arm. The effect size for the technical arm was much smaller (2.1 percentage points higher than the control group) and with 100 kilns per arm, we would not be powered to detect such a small difference. However, 2.1 percentage points is an extremely conservative estimate for a potential effect size. The minimum detectable effect size for 100 kilns per arm at 90% power is 3.56 percentage points. This is half the magnitude of the incentive arm and still relatively conservative, particularly when considering the TOT estimate of 9.22 percentage points among adopters.

Figure 1: Minimum Detectable Effect Sizes for RCT Outcomes



For the CO/CO2 ratio, with 100 kilns per arm, we almost are powered for the more conservative ITT effect attained by the incentive arm but more than sufficiently powered to detect the larger effect size attained by the technical arm. With 100 kilns per arm at 90% power, we are powered to detect an effect size of -0.0064 in the CO/CO2 ratio, while we would need only 65 kilns per arm to detect an effect as large as -0.008, which is what the technical arm attained in the pilot. Somewhat surprisingly, the measured CO/CO2 ratio in the pilot was lower in the technical arm than in the incentive arm. This may simply reflect that the CO/CO2 ratio is a cross-sectional measure that we captured based on data from a few hours in each kiln and so may not accurately reflect the performance over the whole season. Indeed, the first CO/CO2 ratio was measured before the incentive arm was even rolled out. Nevertheless, the calculations suggest that we will have sufficient power to be able to detect changes in CO/CO2 ratio with the interventions.

Similar to the percent of Class-1 bricks, our pilot results suggest kilns assigned to the incentive arm had a much lower specific energy consumption (SEC). While we will not be powered to detect effect sizes as small as what the pilot found in the technical arm, we are powered to detect effect sizes smaller than what the technical arm attained. With 100 kilns per arm at 90% power, we are powered to detect an effect size of -0.065 in SEC, while we would need 70 kilns per arm to detect an effect as large as -0.083, which is the ITT effect for the incentive arm compared to the control group. We summarize the minimum detectable effect sizes for a study with 100 kilns per arm with power of 80% and 90% in Table 5.

It is also of interest to assess the power for detecting differences between the two arms. Although we were not powered in the pilot to statistically detect differences in the exploratory outcomes between the technical and incentive arms, our pilot provides suggestive evidence that kilns assigned to the incentive arm performed better than the technical-only arm, although statistically, we cannot rule out equivalent effects. Using these effect sizes and assuming 100 kilns per arm, we estimated the power we can expect to attain for each outcome, which is presented in Table 5. Given the small differences between the two arms, we are underpowered except for the percent of class-1 bricks, where we estimate having 80% power to detect a difference of 5 percentage points.

Table 5: Minimum Detectable Effect Sizes and Power for 100 kilns per arm

	MDES		**Power b/w treatment arms**
	Power: 0.9	Power: 0.8	
Class-1 (%)	3.56	3.08	0.81
CO/CO2 (ratio)	-0.0064	-0.0056	0.19
SEC (MJ/kg-fired brick)	-0.065	-0.056	0.23

References

Anderson, M. L. (2008). Multiple Inference And Gender Differences In The Effects Of Early Intervention: A Reevaluation Of The Abecedarian, Perry Preschool, And Early Training Projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)*, 57(1):289–300.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

List, J. A., Shaikh, A. M., and Xu, Y. (2016). Multiple Hypothesis Testing In Experimental Economics. *Experimental Economics*, pages 1–21.

Mogstad, M., Torgovitsky, A., and Walters, C. R. (2020). The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables.

Morgan, K. L. and Rubin, D. B. (2012). Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 40(2).

Romano, J. P. and Wolf, M. (2005). Stepwise Multiple Testing as Formalized Data Snooping. *Econometrica*, 73(4):1237–1282.

Romano, J. P. and Wolf, M. (2016). Efficient Computation of Adjusted p-values for Resampling-Based Stepdown Multiple Testing. *Statistics Probability Letters*, 113(C):38–40.