Pre-Analysis Plan: Demand-driven Enforcement of Labor Law in Bangladesh

Laura Boudreau *

Haas School of Business, University of California, Berkeley

December 5, 2017

Abstract

Weak states with poor institutions often do not have the capacity to implement and/or to enforce labor regulations aimed at improving working conditions. Increasingly, private actors have started enforcing labor standards in these countries, but the effects of their interventions on local firms and workers is currently unknown. This paper partners with a set of multinational retail and apparel firms to enforce local labor laws on their suppliers in Bangladesh. Specifically, I design and implement a randomized controlled trial with Bangladeshi garment factories, randomly introducing worker-manager safety committees in supplier establishments. I aim to measure the impacts of this intervention on factories' compliance levels and productivity as well as on workers' welfare. I also explore whether the intervention has an impact on broader equilibrium compliance levels in the sector.

*E-mail: *lboudreau@berkeley.edu*.

Contents

1 Introduction

1.1 Motivation

Global supply chains often extend into weak states with limited social and environmental regulations and with little formal enforcement of existing regulations; well-known examples include the global supply chains for apparel (e.g., Bangladesh, Cambodia, and Myanmar), electronics (e.g., the Democratic Republic of Congo and Zambia), and oil (e.g., Nigeria and Angola). Integration into global supply chains provides benefits to firms and workers in these countries (Tanaka, 2016; Atkin et al., 2016). But lack of regulation and lack of enforcement of existing regulation also carry potentially important but poorly understood social costs. For example, workers often endure dismal working conditions and limited labor rights. Recent experimental research by Blattman and Dercon (2016) finds that industrial work has large, negative health impacts on Ethiopian workers with little to no compensating increases in income.

It is precisely in these poor societies where labor regulation could have the largest impact, but states may not have the capacity to enforce it. In weakly institutionalized environments, incentives provided by private sector actors –for example, large multinational firms – may be able to serve as a substitute for government enforcement. Indeed, multinational firms are increasingly adopting standard setting and enforcement roles in the absence of formal regulation and enforcement (O'Rourke, 2014). Little is known, however, about whether these actions increase targeted firms' compliance and whether they generate net benefits or costs to targeted firms and to their workers.

This research will contribute to the literature on private regulation in global supply chains. The current literature on this topic is confined to descriptive analysis and is unable to answer causal questions. Research suggests that private regulation is not a primary driver of firm compliance and that local context appears to be the main predictor of compliance (Distelhorst et al., 2015; Locke et al., 2013; Locke and Romis, 2010). But to the extent that firms in developing countries respond to their customers' voluntary regulation programs, evidence suggests that compliance is rewarded with increased annual purchasing (Distelhorst and Locke, 2017). This suggests a pathway through which improved compliance may benefit targeted firms.

Furthermore, to the extent that there is overlap between social (and environmental) compliance practices and modern managerial practices, firms pressured to comply with private regulation may benefit from the adoption of these practices. Recent research by Bloom et al (2013) finds that textile firms in India fail to adopt management practices that have important productivity benefits. Several of the management practices that they introduce to firms, in particular operations such as maintenance of machines and of clean production areas, are also aspects of safety compliance. It is reasonable to expect that garment factories in Bangladesh may similarly fail to adopt modern managerial practices due to poor information and to low-skill management staff. This research analyzes whether private enforcement of labor laws results in factories' adoption of productivity-enhancing

managerial practices.

Managers of the firms studied by Bloom et al (2013) report that they failed to adopt many of the modern managerial practices due to lack of knowledge about the practice, or disbelief that adoption of the practice would increase the firm's profits. In Bangladesh, managers' information and capacity is also an important constraint. Fewer than 10% of supervisors in Bangladesh's garments sector have received any formal managerial training, and Macchiavello et al (2015) document that demand for training for these services is very low (in particular for human resource and social compliance training). This research examines the effects of private labor regulation on managers' perceptions of the importance of labor standards on firms' outcomes.

Finally, to my knowledge, nothing is known about the impacts of private regulation efforts on equilibrium compliance levels and production dynamics in the targeted sector. In principle, external standard-setting and enforcement in weak institutional environments could spur a race to the top, in which compliance increases in firms beyond those directly targeted, and the sector benefits from increased access to more and higher margin markets. Conversely, it could spur a race to the bottom, in which more production is moved to firms with worse conditions. Either outcome has important implications for social welfare. I study how private regulation affects the broader industry in which targeted firms compete.

1.2 Research Questions

This research aims to address several important questions. I ask:

1. To what extent can external, private sector enforcement of labor regulation increase compliance in the absence of effective formal enforcement?

2. If private sector enforcement can achieve compliance, what are the effects of buyer-mandated compliance with labor regulation on:
   a. Targeted establishments' productivity?
   b. Targeted establishments' workers?

3. What are the effects of private sector enforcement of labor law on equilibrium compliance levels and production dynamics in the broader industry in which targeted firms compete?

This pre-analysis plan (PAP) concentrates on the first two of these questions. Upon preparation of this PAP, it was unclear whether and to what extent I would be able to incorporate study of spillovers, so the design of this component was in flux. In September, 2017, I obtained the funding for and confirmed the design of the spillovers component (Research Question 3). In October 2017, I amended this PAP in order to incorporate the research design for spillovers. Section 7.2 (Appendix B), pre-specifies my approach.

2 Intervention

I study these questions in the context of Bangladesh's garment sector. Bangladesh plays a critical role in the global apparel supply chain; it is the second largest exporter of garments in the world (behind China) (World Trade Organization, 2015). The country is infamous for limited legal protections for workers and lack of enforcement of the regulations on the books. In a 2011 McKinsey survey of Chief Procurement Officers at major retail and apparel firms, two of the top five major risks to buying from Bangladesh included lack of social compliance and economic and political instability. According to the report: "lssues in social compliance mentioned by some of the top buyers in Bangladesh are the lack of worker education, a remaining risk of subcontracting, lack of law enforcement, and a continued need for developing fair practices and compensation. For example, Bangladesh enacted child labor regulation in 2006, but UNICEF has reported a lack of enforcement."

Several realizations of these risks occurred in a seven-month period over 2012-2013, when several fires and a building collapse killed at least 1,273 workers and injured at least 3,812 workers at exporting factories (Solidarity Center, 2016). In the aftermath of these events, world leaders rebuked the Government of Bangladesh (GoB) for not taking steps to afford internationally recognized worker rights to workers in that country, and the U.S. Government suspended the country's trade benefits under the Generalized System of Preferences (New York Times, 2013). Retail and apparel firms took safety compliance matters into their own hands through the formation of two coalitions of multinational firms. Both coalitions have missions to improve the safety performance and to increase worker empowerment in their Bangladeshi supplier bases. Both coalitions' members agreed to legally-binding, five-year terms.

I collaborate with one of these coalitions, the Alliance for Bangladesh Worker Safety (hereafter, the Alliance), to determine whether private efforts to enforce compliance have effects over and above the government regulation alone. The Alliance's membership includes 29 multinational retail and apparel firms (e.g., Wal-Mart, Gap, Target). The Alliance covers approximately 700 garment factories and 1.21 million workers in Bangladesh. It requires all factories in its supplier base to participate in its building safety audit, building remediation, and worker training and empowerment programs. Failure to comply with one or more of these programs results in suspension from the Alliance's supplier base; as of June 2017, the Alliance had suspended 155 factories. The Alliance is also a member of a Private sector-GoB Factory Closure Panel for cases of imminent danger due to structural integrity, which has fully or partially closed 35 factories.

The Alliance is requiring its suppliers to comply with a 2013 Bangladeshi labor law, and its 2015 rules issuance, that require factories to establish worker-manager Safety Committees (SCs) with equal worker-manager representation and with democratic selection of worker representatives. While the GoB mandated the implementation of SCs, its enforcement of the law is lacking. I study the Alliance's intervention to enforce the law, which is its SC Program. Specifically, the Alliance's SC Program:

1. Establishes SCs through compliant formation processes;
2. Trains SCs' members on their roles and responsibilities, on safety and health topics (S&H), and on leadership and communication skills;
3. Raises awareness of workers and managers on the roles and responsibilities of the SC;
4. Requires SCs to complete a series of activities required by law; and
5. Monitors SCs.

The first four activities occur over a period of 3 months, and the final, monitoring activity continues until six months after the initial engagement.

3 Research design

3.1 Sampling

The target sample size for this study is 80 factories, which is being equally divided between the treatment and the control groups. The sample is being selected through a two-step process. The pool of eligible factories is limited to factories that supply exclusively to the Alliance (and do not supply to the other coalition of multinational firms mentioned above). This is due to a coordination challenge with the other coalition (the eligible pool of factories remains a very relevant population to study).

From this pool of about 300 factories, the Alliance verifies that a factory has a Participation Committee (PC) that has been democratically-elected through free and fair elections. A PC is a worker committee that is responsible for collective bargaining with management. Very few factories have trade unions in Bangladesh, so PCs play a role similar to that of a trade union. For establishments located in Export Processing Zones (EPZs), labor law does not allow the formation of trade unions or PCs, but provides for Workers' Welfare Associations (WWAs) that are supposed to play similar roles. These worker organizations (PCs, trade unions, or WWAs) are responsible for nominating the worker representatives to the SC, and the Alliance requires that they be correctly, democratically formed in order to participate in the SC Program. Once the Alliance verifies that the PC or WWA is compliant, a factory is eligible for the SC Program.

Each time the Alliance verifies a batch of factories, it sends me the list. I randomly assign half of the factories to the treatment group and half to control (after blocking on factory size). So far, 61 of the 80-factory sample size have been randomly assigned. The non-random selection of factories for the evaluation raises an external validity concern. I have access to numerous data sources that will allow me to compare my sample to the broader populations of Alliance-covered and exporting factories in Bangladesh along many relevant characteristics. I am confident that the factory sample is a relevant population that allows me to make generalizable conclusions.

3.1.1 Spillover effects

The main source of spillovers that is a concern is spillovers within firm across factories due to transmission of compliance practices by shared management. In the main evaluation sample, I assume that the nature of these spillovers is such that factories owned by the same firm that are located in the same building or compound are subject to spillovers, while factories in different locations are not subject to spillovers. I will relax this constraint in the spillovers analysis. To deal with the possibility of spillovers across factories owned by the same firm in the same location, I only allow one factory per firm-location to participate in the evaluation group. Pre-randomization, I randomly select one factory per firm-location to participate.

3.1.2 Data collection methods and instruments

I collect several types of data:

1. Factory administrative, safety, and business performance data: Employment, wages, hours, production data, sales, expenditures on material inputs, accidents, and fire histories. SC-related documentation and medical clinic records. The first group is collected once at endline using a retrospective factory questionnaire (at a monthly frequency). The second group is collected three times at baseline, midline, and endline.

2. Stakeholder surveys: Workers, lower-level managers, senior managers, SC worker and management representatives. Surveys are conducted at baseline, midline, and endline.

3. Factory spot-checks: Safety and health compliance spot checks at midline and endline using a safety checklist.

4. Alliance data: Worker Helpline calls, factory building safety audits and remediation progress, monthly supplier lists. I have these data at the monthly frequency.

Teams of five enumerators visit factories at baseline, midline, and endline to administer stakeholder surveys, to collect certain administrative data and verify certain documents, and to conduct spot-checks of safety and health compliance (at midline and endline). One to two members of the team for each visit are professionally trained assessors, and this member(s) implements the document verification and factory spot-checks. For the stakeholder surveys, for workers and lower-level managers, randomly-selected cross sections are selected prior to each visit. For the senior manager survey, the most senior manager present at the factory is surveyed. For the Safety Committee members, the SC President is surveyed (member of management) and two worker members of the SC are randomly selected in advance of the visit.

For most factory administrative, safety, and business performance data, these are collected once at endline using an online factory questionnaire administered with factory management. Most of these data will be collected at the monthly frequency.

3.1.3 Statistical power

The primary level of interest for power calculations in this study is the factory-level. I also conducted power calculations for clustered individual outcomes (e.g., workers and lower-level managers). Ultimately, because the cost of data collection and the logistical challenges are mainly driven by the number of factories, as opposed to the number of surveys per factory, I report the power calculation for the factory sample. I am better powered for individual-level outcomes.

The sample size for the study is 40 treatment factories and 40 control factories. The sample size was determined based on power calculations and resource/logistical constraints. This sample size is sufficient to detect effect sizes of 0.32 standard deviations with 80% power and $\alpha = 0.05$. I believe that this effect size is a moderate effect size for assessment of the SC Program's effects on compliance, a large effect size for effects on productivity, and a moderate effect size for effects on worker wellbeing.

4 Hypotheses

In this section, I state my hypotheses related to my primary outcome variables for factories and for workers. Where hypotheses do not follow directly from the intervention's design, I also provide an explanation for the hypothesis.

H1: The Alliance Safety Committee Program increases compliance with the Bangladesh Labor Act Amendment of 2013 and the associated Safety Committee Rules of September 2015.

H2: The Alliance Safety Committee Program improves factory safety performance.

H3: The Alliance Safety Committee Program may both positively and negatively affect factories' productivity.

Explanation of H3: The SC Program may positively affect factories' productivity through several mechanisms. For example, the SC Program may increase factory safety (e.g., improved machine maintenance, clean and unobstructed passageways, more appropriate use of personal protective equipment (PPE), etc.) in ways that may increase machines' efficiency, reduce worker injuries and occupational diseases, and allow more productive use of factory space. The SC Program may also increase communication flow between workers and managers, which may facilitate more productive organization and implementation of work. Alternatively, the SC Program may act as a tax on productivity, constraining factories' use of capital and labor inputs, and negatively affecting factories' productivity. The SC Program's effects may also be heterogeneous across factories. For example, at factories where managers have the capacity to recognize and to take advantage of SCs' potential benefits, it may increase productivity. At factories where lower capacity managers

are not able to unlock SCs' potential benefits, it may reduce productivity[1].

H4: The Alliance Safety Committee Program may both positively and negatively affect factories' employment and wages.

Explanation of H4: A neoclassical model suggests, barring any increases in labor productivity, that if the SC Program increases factories' labor costs and if SCs are not valued by workers, then factories will lower wages, and employment will fall. If workers value SCs, then wages will fall but employment will fall less or stay the same (depending on whether workers' valuation of SCs is less than or equal to the cost of SCs borne by factories). If one allows factories to have some monopsony power in the labor market, however, then requiring them to establish SCs may actually have positive effects on wages and employment. SCs could also positive effects on wages and employment if they increase labor productivity.

H5: The Alliance Safety Committee Program improves workers' wellbeing and job satisfaction.

Explanation of H5: A direct objective of the SC Program is to improve worker wellbeing through worker empowerment. SCs may also indirectly improve worker wellbeing by improving safety and the healthfulness of the factory environment and through more communication with factory management. If so, then workers may be more satisfied with their work environment and relationships with management, and workers' job satisfaction may increase.

5 Empirical analysis

5.1 Variables of interest

My research questions in Section 1.2 and my hypotheses in the previous section focus on the effects of private sector enforcement of labor regulation on establishments' compliance, productivity, and workers; primary outcomes variables follow directly from these hypotheses. Many mechanisms may underlie effects on these primary outcomes; I therefore also state the numerous secondary outcome variables that I will analyze in order to unpack the effects on my primary outcome variables.

In this section, I list my primary outcome variables, organized by hypothesis. In total, I have six primary outcome variables, some of which are constructed using indices. I discuss construction of indicies and accounting for multiple hypothesis testing in Sub-Section 5.2.5. Next, I list my secondary outcome variables and other outcome variables that I will report in the paper. These secondary and other outcomes are to aid with interpretation of results on primary outcome variables. Finally, I discuss how I will handle reporting in the version of this analysis for an academic audience and the version for industry and policymakers.

[1] I discuss testing for heterogeneous treatment effects in Section 5.2.2.

5.1.1 Primary outcome variables

H1: The Alliance Safety Committee Program increases compliance with the Bangladesh Labor Act Amendment of 2013 and the associated Safety Committee Rules of September 2015.

1. Compliance with Bangladesh SC Regulation (index). Subcategories of compliance:
    - Formation: Formed in accordance with SC Regulation
    - Operations: Operations in accordance with SC Regulation
    - Responsibilities: Spheres of responsibility in accordance with SC Regulation

- *Hypothesized direction of effect:* Positive (increased compliance)

H2: The Alliance Safety Committee Program improves factory safety performance.

2. Factory safety index of SC effectiveness
    - In-factory safety checklist
    - Alliance Correction Action Plan (C.A.P.) (building safety remediation) completion progress
    - Stakeholder awareness of SC
    - Worker safety knowledge

- *Hypothesized direction of effect:* Positive (increased factory safety)

H3: The Alliance Safety Committee Program may both positively and negatively affect factories' productivity.

3. Total Factor Productivity (TFP)[2]

- *Hypothesized direction of effect:* Ambiguous

[2]I will obtain the data to measure TFP for the full period at the RCT's endline. I have piloted the data collection instruments to collect production input and output data and capital data. I aim to collect multiple, different measurements of establishments' capital. I have concerns, however, about whether I will be able to collect high quality data, in particular for measures of capital. If I determine that the data are not sufficiently high quality to use TFP as a primary outcome variable, then I will use labor productivity instead. I will justify this decision if I am necessitated to take this approach.

H4: The Alliance Safety Committee Program may both positively and negatively affect factories' employment and wages.

4. Employment
5. Wages[3]

- *Hypothesized direction of effect:* Ambiguous

H5: The Alliance Safety Committee Program improves workers' wellbeing and job satisfaction.

6. Wellbeing and job satisfaction (index). Subcategories of index:
    - Absenteeism
    - Turnover
    - Mental wellbeing (survey measures)
    - Job satisfaction (survey measures)

- *Hypothesized direction of effect:* Positive (increased wellbeing and job satisfaction)

5.1.2 Secondary outcome variables

Secondary outcome variables for factories:

1. Perceived Safety Committee performance:
    - Stakeholder perception of SC compliance (index)
    - Stakeholder perception of SC effectiveness (index)
2. Labor:
    - Hiring
    - Working hours
    - Non-pecuniary benefits
3. Production:

[3]The Alliance SC Program may have direct effects on labor-related outcome variables and indirect effects through changing the composition of the workforce. I will test whether the SC Program has a compositional effect on treatment factories' workforces. If I find evidence of compositional shifts at treatment factories, then I will also show composition-adjusted regression results for relevant outcome variables.

- Labor efficiency (at garments factories)
- Output quality
- Machine/factory downtime

4. Worker-manager relations (index). Subcategories of index:
    - Strikes and worker unrest
    - Stakeholder perceptions of relations (workers, lower-level managers, senior managers)
5. Supplier-buyer relations (index)

Secondary outcome variables for workers:

1. Worker empowerment (index). Subcategories of index:
    - Safety-related worker empowerment
    - General worker empowerment
2. Awareness of worker organizations (index). Subcategories of index:
    - Awareness of types of worker organizations
    - Awareness of factory's worker organization(s) (Trade union, Participation Committee, Workers' Welfare Association)

Secondary outcome variables for lower-level managers:

1. Perception of worker capabilities and responsiveness to ideas (Index). Subcategories of index:
    - Worker capabilities related to safety
    - General worker capabilities
2. Perception of importance of worker wellbeing, of working conditions, and of worker-management relations for factory's success (index).

5.1.3 Other outcome variables

Accidents, injuries, and reports of working conditions

The ultimate goals of establishing functional Safety Committees include improving workers' health and safety and increasing workers' empowerment related to these issues. Ideally, Primary Outcome #2 would be incidence of worker injuries and illnesses instead of indicators of factory safety performance. I will report effects on worker injury and illness, but I do

not include it as a primary outcome for the study. I take this approach because through increasing worker empowerment, the SC Program may increase workers' reporting of injury and illness. While the overall number of injuries and illnesses at treatment factories may decrease, *reported* injuries and illnesses may increase compared to control factories. The sign of the SC Program's impact on injuries and illness is therefore ambiguous and difficult to interpret.

Similarly, calls to the Alliance's worker helpline may decrease if working conditions improve, if SCs become better equipped to address workers' concerns, or if workers feel more comfortable sharing their concerns with management. Alternatively, calls to the helpline may increase if workers are empowered to report concerns and conditions through this channel.

As with calls related to health and safety, worker calls to the helpline related to other issues (e.g., compensation, freedom of association, treatment by management, etc.) may decrease if conditions improve and/or relations with managers improve. Alternatively, they may increase if workers are empowered to report concerns through the helpline.

a. Worker accidents and illness (index).

   - Medical clinic records: Accidents
   - Medical clinic records: Disease/illness
   - Factory-reported number of fires
   - Factory-reported number of worker injuries
   - Self-reported worker accidents and illness

b. Worker calls to Amader Kotha helpline, safety- and health-related issues.

c. Worker calls to Amader Kotha helpline, non-safety/health -related issues.

Collective action in factories with trade unions or with Workers' Welfare Associations (WWAs)

Factories outside Export Processing Zones (EPZs) are legally required to have Participation Committees (PCs). Trade unions are also legally allowed at factories outside EPZs, although in practice, few factories have trade unions. Inside EPZs, PCs are not required and trade unions are not legally allowed. Instead of trade unions, factories in EPZs are allowed to have Workers' Welfare Associations (WWAs). WWAs are intended to play a similar role as trade unions but are more legally restricted.

Given the limited number of factories in my sample with either trade unions or WWAs, I do not include outcomes related to these organizations as primary outcome variables. I am still uncertain of the total number of factories in the sample that will have either a trade union or a WWA, but it may be too few even to run statistical analysis on the subgroup. I will update

this PAP and will report appropriately depending on the number of factories in this subgroup.

a. Engagement with trade union/WWA

b. Perception of trade union/WWA effectiveness and overall satisfaction

5.1.4 Reporting for a non-academic audience

This RCT is being conducted as an impact evaluation of the Alliance's SC Program. In addition to a research paper for an academic audience, I will prepare an impact evaluation report for the Alliance and results briefs for its partners. The Alliance's and its partners' priorities, in terms of which outcomes variables constitute the primary outcomes of interest, may differ from what academics prioritize. For example, labor efficiency is a Key Performance Indicator (K.P.I.) that factory owners prioritize, while this measure is of less interest to economists than, for example, TFP. While I will construct my outcome variables in the same way in all reporting on the results, the set of outcomes reported for non-academic audiences will vary depending on the target audience. When reporting on the primary outcomes above, I will still also report p-values adjusted for multiple inference, while when reporting on secondary outcomes, I will not show adjusted p-values. For factory owners, for example, the difference between a t-stat of 1.94 or 1.98 is not a critical input into their decision making, while the overall sign and magnitude of the estimated coefficient are more important.

5.2 Treatment effects

Before I move in to regression analysis of the data, I will show baseline balance between the treatment and control groups. I will also show visually the differences in means at treatment and control groups over time.

5.2.1 Basic Specifications

For regression analysis, my basic specifications to identify the SC Program's treatment effects on factory-level outcomes follow best practice in economics for analysis of RCTs with panel data (McKenzie, 2012; Burlig et al., 2017). My approaches include an analysis of covariance (hereafter, ANCOVA) specification and a difference-in-differences (hereafter, DD) specification:

$$Y_{jt} = \alpha + \beta\, T_{jt} + \sum_{t=1}^{t=m} \delta_t + \theta\, Y_{j,t=0} + \mathbf{X}_j'\gamma + \epsilon_{jt} \tag{1}$$

$$Y_{jt} = \alpha + \beta\, T_{jt} + \sum_{t=1}^{t=m} \delta_t + \lambda_j + \epsilon_{jt} \tag{2}$$

where Y_{jt} is the outcome of interest for factory j at time t. T_{jt} is a treatment indicator that is equal to 1 for treatment group factories for all time periods after factory j's baseline

survey and 0 otherwise.

The δ_t are time dummies, where the omitted dummy is the first time period; m is the final period, the value of which will vary depending on data type. In Equation (1), $Y_{j,t=0}$ is the baseline value of the outcome variable, which is included in order to improve efficiency (due to autocorrelation in outcomes across time). Also in Equation (1), $\mathbf{X}_j$ is a vector of controls for baseline stratification variables, including factory size and factory batch[4]. In Equation (2), the λ_j are factory fixed effects, which absorb baseline controls. In both equations, ϵ_{jt} is the residual, which is clustered by factory. β identifies the SC Program's treatment effect on treatment factories relative to control factories.

In the main analysis, I will use either the ANCOVA or the DD approach, whichever provides me with better power to detect treatment effects. In the power simulations conducted with some available factory variables, in short panels, ANCOVA provided greater power than DD for available variables; in long panels, however, whether ANCOVA performs better or worse than DD depended on the outcome variable[5]. Because I do not have access to pre-existing panel data for my primary outcome variables, it is very difficult to predict which strategy will ultimately provide greater power to detect effects[6]. Hence, in the main paper, I will display whichever approach provides greater power for a greater number of primary outcome variables.

When analyzing individual-level outcomes, which are clustered at factory-level, I will similarly use one of the following specifications:

$$Y_{ijt} = \alpha + \beta\, T_{jt} + \sum_{t=1}^{t=m} \delta_t + \theta\, \bar{Y}_{j,t=0} + \mathbf{X}_j'\gamma + \bar{\mathbf{Z}}_j'\eta + +\epsilon_{ijt} \tag{3}$$

$$Y_{ijt} = \alpha + \beta\, Treatment_{jt} + \sum_{t=1}^{t=m} \delta_t + \lambda_j + \epsilon_{ijt} \tag{4}$$

In this specification, the i indicates that the unit of observation is an individual. The research design uses a repeated cross section. In Equation (3), $\bar{Y}_{j,t=0}$ is the mean value of the outcome variable for individuals at factory j at $t = 0$. $\bar{\mathbf{Z}}_j'$ is a vector of mean values of individual characteristics at factory j at $t = 0$. These include sex, age, educational attainment, tenure, and work experience. As before, standard errors will be clustered at the factory-level.

[4]Specifically, factories are stratified using a dummy variable for having greater than or equal to 3000 employees and using dummy variables for "readiness for SC Program," or when the factory met Alliance prerequisites for participation in the SC Program. Ex ante, I anticipate that these variables will be associated with factories' potential outcomes, so stratifying on them will increase the precision of my estimates. Ex post, if they are not associated with factories' outcomes, then I reduce the power of my tests if I include them in my regressions (Imbens, 2011); in this case, I will not include these variables in my regressions.

[5]For certain variables, such as compliance and worker wellbeing, I will have three rounds of data; for others, such as the human resource and production data, I will have monthly data.

[6]See Burlig et al. (2017) for discussion of power using ANCOVA versus DD designs in experiments using panel data.

5.2.2 Heterogeneous effects

I will test whether the treatment effect varies with pre-determined factory characteristics. I will analyze the following dimensions of heterogeneity:

1. Factory location (EPZ vs. non-EPZ)
2. Factory size[7]
3. Baseline factory compliance with SC regulation (index)
4. Baseline management practices (index)

5.2.3 Temporal dynamics of treatment effects

In my baseline specification, β is the average treatment effect over all post-treatment periods. If the treatment's impact varies over time, then this specification will mask this dynamic. Since I am interested in understanding such dynamics, I will also split the treatment indicator into two separate indicators for the midline (t=*4* months) and the endline (t=*8* months). For outcomes with monthly data, I will group the data to be consistent with this approach. I will loose some precision in this analysis, but comparing the separately-estimated coefficients from this analysis to the estimated coefficient on T_{jt} from (1) will provide insight into whether the the treatment effects exhibit temporal dynamics. My specification is as follows[8]:

$$Y_{jt} = \alpha + \beta_1 \, Treatment_{j,t=1} + \beta_2 \, Treatment_{j,t=2} + \theta \, Y_{j,t=0} + \sum_{t=1}^{t=2} \delta_t + \mathbf{X}_j' \gamma + \epsilon_{jt} \qquad (5)$$

$$Y_{jt} = \alpha + \beta_1 \, Treatment_{j,t=1} + \beta_2 \, Treatment_{j,t=2} + \sum_{t=1}^{t=2} \delta_t + \lambda_j + \epsilon_{jt} \qquad (6)$$

5.2.4 Standard error adjustment

I will cluster standard errors at the factory-level in all regressions.

[7] If I find evidence of heterogeneous effects by factory size, it may suggest that my estimates of Average Treatment Effects (ATEs) for lower-level factory managers are biased (Middleton and Aronow, 2011). They may be biased because the cluster sizes for lower-level managers varies with factory size – small factories have few mangers and smaller cluster sizes than large factories – and factory size may correlated with managers' potential outcomes. In this case, I will show estimates of ATEs for lower-level managers following Middleton and Aronow (2011).

[8] Note that for ease of exposition, I show subscripts for a regression with two post-treatment rounds of observations.

5.2.5 Adjustments for multiple inferences

I test several hypotheses related to how the Alliance's SC Program impacts factories. I will also show the p-values for the coefficients of my primary outcome variables adjusted for multiple statistical inference. I also construct several primary, secondary, and other variables using indices. To construct these indices, I will follow the procedure from Anderson (2008). For primary outcome variables that I construct using indices, for all variables in the index, I will also show the individual t-tests and their p-values as well as False Discovery Rate adjusted p-values.

For index variables, I am still in the process of constructing the indices. I will update this PAP by attaching the indices as an appendix.

6 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.

Atkin, D., A. Khandelwal, and A. Osman (2016). *Exporting and Firm Performance: Evidence from a Randomized Experiment*. Working Paper.

Baird, S., A. Bohren, C. McIntosh, and B. Özler (2016). *Designing Experiments to Measure Spillover Effects*. World Bank Policy Research Working Paper 6824.

Blattman, C., and S. Dercon (2016). *Occupational Choice in Early Industrializing Societies: Experimental Evidence on the Income and Health Effects of Industrial and Entrepreneurial Work*. Working Paper.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, J. Roberts (2013). *Does Management Matter? Evidence from India. Quarterly Journal of Economics* 128(1), 1-51.

Burlig, F., Preonas, L., and Woerman, M. (2017). *Panel Data and Experimental Design*. Working Paper, Energy Institute at Haas, University of California, Berkeley.

Distelhorst, G., R. Locke, T. Pal, and H. Samel (2015). Production goes global, compliance stays local: Private regulation in the global electronics industry. *Regulation & Governance* 9(3), 224-242.

Imbens, G. W. (2011). *Experimental design for unit and cluster randomized trials*. Working Paper, Harvard University.

Locke, R.M., B.A. Rissing, and T. Pal (2013). Complements or Substitutes? Private Codes, State Regulation and the Enforcement of Labour Standards in Global Supply Chains. *British Journal of Industrial Relations*, 51(3), 519-552.

Locke, R.M. and M. Romis (2010). The promise and perils of private voluntary regulation: Labor standards and work organization in two Mexican garment factories. *Review of International Political Economy*, 17(1), 45-74.

Macchiavello, R., A. Rabbani, and C. Woodruff (2015). The Market for Training Services: A Demand Experiment with Bangladeshi Garment Factories. *American Economic Review: Papers & Proceedings*, 105(5), 300-304.

McKenzie, David (2012). Beyond Baseline and Follow-up: The Case for More T in Experiments. *Journal of Development Economics*, 99(2): 210-221.

Middleton, J. A., and Aronow, P. M. (2015). Unbiased Estimation of the Average Treatment

Effect in Cluster-Randomized Experiments. *Statistics, Politics and Policy*, 6(1-2), 39-75.

ORourke, D. (2014). The science of sustainable supply chains. *Science* 344(6188), 1124-1127.

Tanaka, M. (2016). *Exporting sweatshops? Evidence from Myanmar*. Working Paper.

7 Appendices

7.1 Appendix A: List of revisions to Pre-Analysis Plan

6-27-2017:

1. Subsection 5.1.2, Secondary outcome variables for workers: *Perception of worker selection to worker organizations* removed from second secondary outcome variable for workers (*Awareness of worker organizations (index)*).

2. Subsection 5.1.3, Worker accidents and illness outcome variable. Corrected typo: *Self-reported worker accidents and injuries* originally input as distinct outcome from the index variable, but actually should have been included on the list as the fifth bullet point. Corrected in this version of the PAP.

12-05-2017:

1. Section 7.2: Added Appendix B to pre-specify my approach to analysis of cross-establishment spillovers.

2. Primary outcome index variables: Updated to include endline variable numbers, and some minor modifications made to the index variables.

7.2 Appendix B: Cross-factory spillovers

If the Alliance's SC Program has effects on treatment factories' compliance with Bangladeshi labor law, a key question is whether the effects spillover to other factories. I analyze several potential mechanisms for spillovers that I identify as plausible based on conversations with the Alliance and on my other research with Bangladeshi garment workers. While these are not the only possible mechanisms that may be at work, they are good candidates and are viable to study.

The first group of potential mechanisms generates spillovers via changes in compliance practices across factories within a firm. If the SC Program has effects on treatment factories, there may be spillovers across factories owned by the same firm due to corporate management updating policies or making decisions that affect all of the firm's factories. For example, firms may learn about the benefits and costs of SCs and change corporate requirements for SCs or reallocate production across factories. Alternatively, buyers may pressure firms to improve SC performance in their factories that have not participated in the SC Program. Another channel within the firm through which spillovers may occur is communication among factory managers. For example, factory managers may learn about benefits and costs of SCs from treatment factory managers and change their factory's SC practices.

A different potential mechanism for spillovers is information sharing across workers who work at factories nearby to each other. I will document the plausibility of this mechanism by assessing the extent to which workers at SC Program treatment and control factories report sharing information related to worker committees and safety conditions with workers outside their factory. Given my financial constraints and my means of accessing workers, however, I am not well situated to study spillovers via information sharing by workers. So, I focus on the cross-factory within a firm mechanisms discussed above.

7.2.1 Research design

Multi-factory firms are very common in Bangladesh's garments' sector. At the time of writing, I am developing the data for the Alliance's broader supplier base, but among the factories in the RCT, about 70% are owned by firms with multiple garments factories. Furthermore, the Alliance's coverage of the sector is great enough that many of the factories owned by these firms supply to Alliance members, which provides an entry point for data collection. These features enable me to test for the presence of spillovers across factories owned by the same firm.

The RCT's design entails random assignment of factories to either treatment or control status. As discussed in Section 3.1, in cases where more than one factory owned by a single firm are located at the same location, one factory is randomly selected to participate in the RCT. In this context, "same location" is defined as either the same building or the same compound (where a compound is a plot of land housing multiple factories at the same

address). All Alliance-covered factories at the location share the assignment status of the factory selected to participate in the RCT for the duration of the evaluation period. This research design accommodates the possibility of spillovers across factories under shared ownership in the same location. It assumes, however, that there are not spillovers across factories under shared ownership in different locations.

In this analysis of spillovers, I relax this assumption within firms. I allow spillovers across all factories owned by the same firm. I maintain the stable unit treatment value assumption (SUTVA) across firms. This ensures that pure control firms provide a valid counterfactual for treated firms and that my estimates are causal.[9] I am collecting information about information sharing across firms that will allow me to assess the plausibility of this assumption.

The data collection for the spillovers analysis occurs at endline for the first factory in the firm that is participating in the RCT. Specifically, factories will be invited to participate in the spillovers data collection within one week following the endline completion for the factory participating in the RCT.

Sampling

Firms are eligible to participate in the spillovers analysis if they are multi-factory, have at least one factory participating in the RCT, and have at least one eligible factory. Alliance-covered factories are eligible to participate in the spillovers analysis if they are not eliminated by any of the following criteria:

- Owned by a firm with no factories participating in the RCT;
- Share the same location with a factory participating in the RCT;
- Have participated in the Accord's Safety Committee Program;
- Have participated in the Alliance's pilot of its Safety Committee Program;
- Have been suspended by the Alliance.

From among the eligible firms, I select factories using the following rules. For firms owning 1-4 eligible factories, all eligible factories are invited participate in the data collection. For firms owning 5 or more eligible factories, four factories are randomly selected to be invited to participate in the data collection. I impose this cap in order to prevent the risk of firms being overwhelmed by the participation requests (in many cases, one contact person receives an invitation for multiple factories).

[9] If I find evidence of the presence of within firm, cross-factory spillovers, this means that my main effect estimates are biased. In this case, I will consider how this impacts my estimates in the discussion of the main results.

7.2.2 Hypotheses

I focus on the intra-firm mechanisms for spillovers, which entail transfer of compliance practices across factories within a firm. To assess the plausibility of transfers via changes in corporate policy, I will collect information about to what extent corporate versus factory management set safety-related, and in particular, SC-related policies. I will also collect whether corporate-level policies change in response to treatment factories' participation in the SC Program. To assess the plausibility of transfers via learning from treatment factories' managers, I will ask about information sharing across managers and to what extent managers at different factories report changing their practices based on what they learn through these exchanges. I will also ask how managers learn about the SC Program (if they are aware of the program), including learning about it from buyers.

Next, I will test whether managers at firms with more factories participating in the Alliance SC Program are more aware of it. I hypothesize that greater treatment saturation of the firm (i.e., greater proportion of factories randomly assigned to participating in the Alliance SC Program) will result in managers at non-treatment factories within the same firm having greater awareness of the SC Program.

Spillovers H1: At firms with a greater proportion of their garments factories participating in the Alliance SC Program, non-treatment garments factory managers will be more aware of the Alliance SC Program.

Ultimately, I would like to test for spillover effects on SC compliance. I am limited, however, by my sample size for the spillovers group and by the lack of preexisting studies or data to determine plausible spillover effect sizes. At present, I will have 50 factories in the spillovers group and 41 evaluation group factories under shared ownership with these factories for a total of 91 factories that may potentially be included in the analysis. The final number of factories in the analysis could increase or decrease due to fluctuations in the Alliance's supplier base and Accord intervention in potential spillovers factories. Given my uncertainty, this analysis will be an exploratory analysis that I will implement if I have a sufficient sample size and a large enough main effect size that I am plausibly powered to detect spillover effects.

As discussed in Section 1.1, in principle, if the SC Program results in spillover effects on SC compliance, they could be positive or negative in nature. I hypothesize that the direction of the effect will depend on the treatment's actual or perceived (by management) impact on productivity. If the SC Program increases factory productivity (or has a null effect on productivity and other positive effects, such as increased worker job satisfaction), management may transfer compliance practices learned through the SC Program to other factories within the firm.

Alternatively, if the SC Program decreases productivity, factory management may respond in ways that decrease SC compliance at other factories within the firm. First, management could try to offset decreases in productivity at treatment factories by shifting production to other factories, and the increased production pressure could result in lower safety-related

compliance. Second, management could intentionally reduce compliance at less-closely monitored factories because they have learned about the negative impact of these practices on productivity, or these impacts become more salient.

Spillovers H2: The Alliance Safety Committee Program may result in positive or negative spillover effects on compliance across establishments owned by the same firm.

7.2.3 Empirical analysis

Primary outcome variables

Spillovers H1: At firms with a greater proportion of their garments factories participating in the Alliance SC Program, non-treatment factory managers will be more aware of the Alliance SC Program.

1. Aware of the SC Program.
2. Able to name a requirement of the SC Program.

- *Hypothesized direction of effect:* Positive (increased awareness)

Possible exploratory outcome variable

Spillovers H2: The Alliance Safety Committee Program may result in positive or negative spillover effects on compliance across establishments owned by the same firm.

1. Basic compliance with Bangladesh SC Regulation (index).[10] Subcategories of compliance:
    - Formation: Formed in accordance with SC Regulation
    - Operations: Operations in accordance with SC Regulation
    - Responsibilities: Spheres of responsibility in accordance with SC Regulation

- *Hypothesized direction of effect:* Ambiguous

Treatment effects

I will use pre-evaluation data to test for balance. At this time, I uncertain what data will be available for all spillover factories, but aim to use Alliance administrative data (e.g., calls to its worker helpline) and other available data (e.g., number of employees, building safety audit results, etc.) for balance tests.

[10]The basic index includes a subset of outcomes from the main RCT index of compliance that will be collected for spillover factories.

I assess the plausibility of possible mechanisms by characterizing to what extent corporate versus factory management is responsible for SC-related policies. I will also assess the extent to which managers report policy changes in response to factories' participation in the SC Program. I will also characterize the extent of information sharing by managers of different factories within and across firms.

Specifications

For my primary outcome variables, I will estimate spillover effects by imposing a linear model and estimating how spillover effects vary with a firm's treatment intensity.

$$Y_{jc} = \alpha + \beta_1 S_{jc} + \beta_2 (S_{jc} * \pi_c) + \mathbf{X}'_{jc} + \mathbf{Z}'_c \eta + \epsilon_{jc} \tag{7}$$

where Y_{jc} is the outcome of interest for factory j in firm c. S_{jc} is an indicator that is equal to 1 for non-treated factories within firms with at least one factory in the treatment group. S_{jc} is equal to 0 for non-treated factories within firms with one or more factories in the control group, but no factories in the treatment group (pure control firms). π_c is the proportion of a firms' factories that are treated. $\mathbf{X}_{jc}$ is a vector of pre-determined control variables, and $\mathbf{Z}_c$ is a vector of firm control variables (e.g., firm size, proportion of factories that participated in the Alliance's SC Program pilot, proportion of factories that participated in the Accord's SC Program, etc.). ϵ_{jc} is the residual, which is clustered by firm.

There is variation in firms' treatment saturation, π. Among firms where not all factories are treated, π, ranges from 0 to 0.67. This variation allows me to estimate β_2, the slope of the relationship between treatment saturation and the spillover effect. The intercept, β_1, estimates spillover effects at saturation equal to zero; there should be no spillover effects on factories if the saturation of treatment is equal to zero, so this estimate serves as a hypothesis test for the linearity of the spillover relationship. [11]

If I implement the exploratory analysis for Spillovers H2, I will modify the equation above as follows:

$$Y_{jc} = \alpha + \beta_1 S_{jc} + \beta_2 T_{jc} + \beta_3 (S_{jc} * \pi_c) + \beta_4 (T_{jc} * \pi_c) + \mathbf{X}'_{jc} + \mathbf{Z}'_c \eta + \epsilon_{jc} \tag{8}$$

where T_{jc} is an indicator for a treatment factory. Control factories will also be included in the analysis. Control factories at firms with at least one treatment factory will be in the group of factories with $S_{jc} = 1$.

In Equation 8, β_3 is the relationship between a firm's treatment saturation and spillovers on non-treated factories, and β_4 is the relationship between a firm's treatment saturation and spillovers on treated factories. β_2 provides the same test of linearity of the spillover relationship as in Equation 7. β_1 is the treatment on the uniquely treated, which is the intention to treat effect on a treatment factory in a firm where only one factory is treated.

[11] See Baird et al., 2014, for a detailed discussion for estimation of spillover effects.