Multinational enforcement of labor law: Experimental evidence from Bangladesh's apparel sector

Laura Boudreau*

January 29, 2021

Abstract

Western stakeholders are increasingly demanding that multinationals sourcing from developing countries be accountable for working conditions upstream in their supply chains. In response, many multinationals privately enforce labor standards in these countries, but the effects of their interventions on local firms and workers are unknown. I partnered with 29 multinational retail and apparel firms to enforce local labor laws on their suppliers in Bangladesh. I implemented a field experiment with 84 garment factories, randomly enforcing a mandate for safety committees. The intervention increases compliance with the law and improves measures of safety. My findings are consistent with a model of imperfect monitoring in which MNCs provide positive penalties for noncompliance. These improvements do not appear to come at significant costs to suppliers in terms of efficiency. Factories with better managerial practices drive the improvements, while those with poor practices do not improve, and in these factories, workers' job satisfaction declines. JEL Codes: F61, J53, J81, L14, O12, O14

Keywords: Labor Regulation, Enforcement, Global supply chains, Corporate Social Responsibility, Private regulation, Working Conditions, Worker voice, Safety, Garment Industry, Bangladesh

*Columbia University & CEPR (l.boudreau@columbia.edu). I thank Noam Yuchtman, Reed Walker, Ernesto Dal Bó, Paul Gertler, and Ted Miguel for their guidance and encouragement. This research has benefited from comments and suggestions from Matt Amengual, Marianne Bertrand, Nick Bloom, Wouter Dessein, Greg Distelhorst, Fred Finan, John de Figueiredo, Jonas Hjort, Amit Khandelwhal, David Levine, Rocco Macchiavello, Jeremy Magruder, Felix Oberholzer-Gee, Raul Sanchez de la Sierra, Daniela Scur, Abu Shonchoy, Mike Toffel, and numerous seminar and conference participants. This project is funded by the Alliance for Bangladesh Worker Safety, the Private Enterprise Development in Low-income Countries (PEDL) Initiative, the Weiss Family Program Fund for research in Development Economics, the International Growth Centre (IGC), and the Abdul Latif Jameel Poverty Action Lab (J-PAL) Governance Initiative. The Sylff Research Abroad Program also contributed travel funding.

1 Introduction

Over the past four decades, developing countries have become increasingly incorporated into global value chains (GVCs), which has raised economic growth and reduced poverty (World Bank, 2020). Despite these gains, poor working conditions and labor rights commonly persist (ILO, 2016). In light of weak local enforcement capacity and corruption (La Porta et al., 1999, Fisman and Wang, 2015, Amirapu and Gechter, 2020), Western stakeholder are increasingly demanding accountability from multinational corporations (MNCs) that source from developing countries. In response, MNCs commonly implement "Corporate Social Responsibility" (CSR) programs that claim to privately enforce local labor laws on their suppliers. These efforts increasingly take the form of participation in multi-firm and multi-stakeholder enforcement initiatives, which now span the globe as well as many GVCs.

MNCs' voluntary adoption of private enforcement suggests that they may have incentives to prevent violations (Kitzmueller and Shimshack, 2012, O'Rourke, 2014), at least those that could pose reputational risks (Macchiavello and Morjaria, 2015, Bai, Gazze and Wang, 2019). On the other hand, increased protections for workers may erode suppliers' competitiveness (Botero et al., 2004, Besley and Burgess, 2004) and punishing noncompliance may be costly to MNCs, which suggests that their promises may not be credible (Besley and Ghatak, 2007). Finally, even if MNCs are motivated to improve labor standards, it's unclear whether their suppliers are capable of adopting better practices.

This paper provides the first experimental evidence on the effects of private enforcement of labor law by MNCs. To do so, I partnered with a group of multinational retail and apparel firms, the Alliance for Bangladesh Worker Safety (hereafter, the Alliance), that expressed a desire to improve its Bangladeshi suppliers' safety performance.[1] The Alliance's membership counted 29 multinational retail and apparel firms, including the two largest global retailers and eight others in the top 100 (e.g., Walmart, Costco, Target); together, these 10 firms accounted for 23% of the top 100 global retailers' revenues (Deloitte, 2015).[2] In response to weak local freedom of association (FOA) rights and

[1] As per the Alliance Members' Agreement, the Alliance ceased operations on December 31, 2018. Upon its exit, many Alliance members supported the formation of Nirapon, an organization tasked with a similar set of safety oversight functions, to continue safety enforcement. 22 out of 29 Alliance Members joined Nirapon.

[2] Alliance Members: Ariela and Associates International LLC; Bon Worth; Canadian Tire Corporation, Limited; Carter's Inc.; The Children's Place Retail Stores Inc.; Costco Wholesale Corporation; Fruit of the Loom, Inc.; Gap Inc.; Giant Tiger; Hudson's Bay Company; IFG Corp.; Intradeco Apparel; J.C. Penney Company Inc.; Jordache Enterprises, Inc.; The Just Group; Kate Spade & Company; Kohl's Department Stores; L. L. Bean Inc.; M. Hidary & Company Inc.; Macy's; Nordstrom; One Jeanswear Group; Public Clothing Company; Sears Holdings Corporation; Target Corporation; The Warehouse; VF Corporation;

power imbalances between workers and their employers, the Alliance stated an aim to empower workers to take an active role in their safety and to be able to report unsafe conditions without risk of retaliation (Alliance, 2013). To do so, it committed to enforce Bangladesh's recent mandate for safety committees (SCs), which are joint worker-manager bodies tasked with helping to create and to maintain a safe workplace. Factory owners fiercely opposed the mandate.

In collaboration with the Alliance, I implemented a randomized controlled trial (RCT) in which I randomly assigned supplier factories to the Alliance's enforcement of the SC law. The Alliance's intervention entailed six months of intensive monitoring of factories' adherence to the SC law through mandated reporting of SCs' activities and monitoring via email, phone calls, and onsite visits.[3] I estimate the intervention's effects on suppliers' compliance with the law and on indicators of safety. I also assess its effects on factories' labor productivity, wages, and employment and on their workers' well-being.

To gain greater insight into the credibility of the Alliance's enforcement regime, I develop a theoretical framework for suppliers' compliance decisions under private enforcement. The model generates different testable predictions depending on Alliance members' willingness to penalize noncompliance and to exert monitoring effort. The fact that MNCs' true motivation may be for their suppliers to achieve formal compliance, rather than meaningful safety improvements, creates the possibility for suppliers to engage in box ticking: Cheap compliance that does not improve safety. For example, suppliers may opt to comply with requirements such as maintaining a list of SC members (box ticking), while foregoing those such as conducting regular factory risk assessment (costlier but effective compliance). The framework generates predictions for suppliers' baseline levels of cheap and costlier but effective compliance and their responses to the SC Program.

The RCT was implemented with 84 garment and garment-related factories that collectively employ nearly 92,000 workers; it was conducted as part of the SC Program roll-out over 2017-2018. I randomly assigned 41 factories to immediate participation (treatment) and 43 factories to deferred participation approximately 11 months later (control).[4] The research team made three full-day visits to factories: A pre-intervention baseline and two post-intervention rounds about 5 and 9 months later, respectively. For treatment factories, the second visit occurred toward the end of the 6-month enforcement intervention. I also implemented a retrospective questionnaire to collect factories' business-related data. Fi-

Walmart Stores, Inc.; and YM Inc.

[3]Post-Alliance, the organization Nirapon is implementing a safety committee program that is based on the Alliance's Safety Committee Program.

[4]Factories were not aware of their experimental status. Due to logistical constraints, the Alliance rolls out all of its programs in stages, so this design aligns with the Alliance's standard operating procedures.

nally, the Alliance provided its administrative records. These datasets are unique in their comprehensiveness and depth. I analyze them according to a pre-analysis plan (PAP), which is registered on the American Economic Association's Social Science Registry.

I find that the MNCs' enforcement program significantly increases factories' compliance with Bangladesh's SC law, which I measure using an index of compliance outcomes. The intervention improves compliance by 0.20 standard deviations (sds) on average. The improvement is driven by factories' engagement in costlier but more effective requirements, such factory risk assessment, as opposed to box ticking. While factories' substantially improve effective compliance, however, they do not achieve full compliance. In my framework, these results are consistent with MNCs' providing positive but relatively smaller penalties for noncompliance. Evidently, MNC enforcement can cause factories to go beyond box ticking and to increase costly, effective compliance with the law, but MNCs may opt *not* to set penalties high enough so as to induce full compliance.

The increase in effective compliance translates into a statistically significant improvement in an index of factory safety indicators (0.14 sds), which is mostly driven by an improvement in treatment factories' performance on an independent spotcheck of safety conditions. Medical clinic records available for 62 factories also document that the proportion of the workforce seeking medical care declines by between 15-16%. MNCs' enforcement interventions can meaningfully improve working conditions.

There is no evidence of adverse effects on supplier competitiveness, including labor productivity, wages, and employment. The estimated treatment effects on labor productivity are positive (not statistically significant), and with 95% confidence, are *not* more negative than -3.3 to -4%. Together, the results suggest that the intervention improved safety without negative effects on suppliers' competitiveness. In light of the null results for these outcomes, I also report ex post Minimum Detectable Effect Sizes (MDEs).

Finally, I explore whether providing firms with incentives to comply with the SC law is sufficient to increase compliance or whether firms' capacity to respond to incentives also matters. The existing research on improving adherence to regulation in developing countries focuses on strengthening state-supplied enforcement in order to increase firms' incentives for compliance (Duflo et al., 2013, 2014). There has been little to no consideration, however, of whether the capacity of the private sector also constrains regulatory efficacy. I test whether the intervention's effects depend on firms' capacity using a measure of their management practices.

I find that factories' baseline managerial practices are an important factor in deter-

mining the enforcement intervention's effects.[5] The treatment has large, positive effects on effective compliance and on safety indicators in factories with better baseline managerial practices. In contrast, factories with worse practices do not significantly improve. The treatment effect heterogeneity is also present in the medical clinic results. These results suggest that there may be complementarity between labor regulation and managerial practices. Increasing compliance may depend not only on providing firms with appropriate incentives, but also on their capacity to respond to these incentives.

Further, the enforcement intervention has a negative effect on indicators of workers' job satisfaction in poorly-managed factories, measured using an index of survey and revealed preference measures. I provide suggestive evidence of a mechanism in which the intervention raises workers' expectations about what SCs will deliver, and in poorly-managed factories, these expectations are not met. This mechanism would be consistent with experimental evidence on low-wage workers' response to an upgrade in employer-provided housing from Adhvaryu, Nyshadham and Xu (2018). Objective measures of housing quality improve, but do not meet workers' expectations, and turnover increases.

This research makes four primary contributions. First, it contributes to the literature on labor regulation and economic development, and in particular, their interaction with GVCs. Studies have found that heavier *de jure* labor regulation is associated with worse economic performance and adverse consequences for workers (Fishback and Kantor, 1996, Botero et al., 2004, Besley and Burgess, 2004, Aghion et al., 2008). We also know, though, that weak state capacity and political capture by elites results in socially suboptimal enforcement in developing countries (Fisman and Wang, 2015, Amirapu and Gechter, 2020). Scholars have raised GVCs as a possible mechanism to bring about improved enforcement. Harrison and Scorse (2010) show that anti-sweatshop campaigns led the Indonesian government to raise minimum wages, which caused large real wage increases with some costs for firms but no significant effects on employment. Tanaka (forthcoming) provides evidence of trade-induced social upgrading among firms in Myanmar. This is the first study, however, to test the potential for MNC enforcement in a context where state enforcement is lacking. My findings demonstrate that private enforcement can improve compliance. I further contribute by identifying enforcement's causal effects on labor productivity and workers' well-being. Finally, my results suggest an under-explored constraint on regulatory efficacy: Managerial capacity of the private sector.

Second, it contributes to a burgeoning literature on the economics of CSR. Economists have long espoused Friedman (1970)'s view that markets should produce private goods

[5]For certain variables, due to power limitations, I am unable to reject that the estimated treatment effects are different for the two groups.

and governments should provide public goods and correct failures. Recent theoretical and empirical work, however, highlights two reasons why this dichotomy may blur. First, governments, particularly in developing countries, frequently fail to fulfill their aforementioned roles; further, governments' jurisdiction is limited to their territories, and they are often constrained in their ability to police production abroad (Bénabou and Tirole, 2010). Second, consumers, shareholders, and workers have social and ethical motivations and often value production that occurs socially and environmentally responsible ways (Dragusanu, Giovannucci and Nunn, 2014, Hainmueller, Hiscox and Sequeira, 2015, Burbano, 2016, Hart and Zingales, 2017). The existing economic literature on CSR primarily provides the bases for its existence and desirability. This paper joins Dragusanu and Nunn (2018) and Macchiavello and Miquel-Florensa (2019), both of which study the welfare effects of CSR programs on coffee farmers, in beginning to build a body of empirical evidence on the efficacy of CSR. I provide the first experimental evidence that firms' CSR initiatives can generate public goods/curtail public bads.[6]

Third, this paper contributes to the literature on collective worker voice and intra-firm institutions, including occupational safety and health (OSH) committees. The empirical literature on this topic has generally suffered from selection bias, and until recently, available causal evidence has largely been limited to marginal firms (DiNardo and Lee, 2004, Lee and Mas, 2012, Yao and Zhong, 2013, Jäger, Schoefer and Heining, 2020). A much smaller literature examines correlations among the presence and features of OSH committees and injury rates or stakeholder satisfaction with them (see Yassi et al. (2013) for a review). My contribution is to randomize enforcement of worker-manager OSH committees to provide causal evidence of their effects on factories' and workers' outcomes.

Finally, this paper contributes to a nascent literature on the role of *private* sector capacity in determining the efficacy of *public* regulation. Existing research focuses on determinants of public sector capacity – the selection of public servants (Dal Bó, Finan and Rossi, 2013), their financial incentives (Ashraf, Bandiera and Jack, 2014), and their human capital and public sector motivation (Dal Bó, Finan and Rossi, 2013, Callen et al., 2018, Ashraf et al., forthcoming). In contrast, my analysis of heterogeneous treatment effects by management practices shows that the effects of regulation also depend on private sector capacity. Regulatory compliance is a function of both firms' incentives for compliance

[6] A literature spanning political science and management science asks related questions. This literature is largely skeptical, but as more causal evidence emerges, it is updating its view. In particular, Amengual and Distelhorst (2019), who also provide an excellent summary of the literature, use a regression discontinuity design to study Gap Inc's supplier code of conduct for labor. They find that a failing audit grade only improves compliance if coupled with the threat of a reduction in Gap's orders.

and their managerial capacity to respond to these incentives.[7]

The remainder of this paper is organized as follows: Section 2 describes the context, the Alliance, and the SC Program. Section 3 presents a conceptual framework that generates different testable predictions for suppliers' responses to the SC Program depending on Alliance members' willingness to penalize noncompliance and to exert monitoring effort. Section 4 presents the research design. Section 5 presents the results and discusses the extent to which they generalize to other settings. Section 6 concludes.

2 Background

2.1 Private enforcement in GVCs

While there are no formal statistics on what share of production in GVCs is subject to private enforcement, there is reason to believe that it is substantial. Private enforcement efforts take many forms, including supplier codes of conduct, certification requirements, and, increasingly, multi-firm and multi-stakeholder initiatives (hereafter, "coalition-based approaches") (de Bakker, Rasche and Ponte, 2019).[8] Coalition-based approaches are organized both at the global- and country- or regional-level and target both broad categories of social and environmental standards (e.g., the Responsible Business Alliance monitors social and environmental standards in electronics, retail, auto and toy GVCs) and specific issues in a given GVC (e.g., the Responsible Mica Initiative aims to reduce child labor in mica mining in India).[9] Despite MNCs' widespread adoption of private enforcement, there is a great deal of skepticism about their influence on labor standards (Locke, 2013).[10]

2.2 Bangladesh's apparel sector

Bangladesh plays a critical role in the global apparel supply chain. It is the second largest exporter of clothing in the world behind China (World Trade Organization, 2017). MNCs

[7] Almunia et al. (2020) find that firms' capacity, broadly defined, also plays a role in determining tax compliance, with some firms disadvantageously over-reporting their taxes when enforcement is low.

[8] See Section 5.5 for a discussion of some of the rationales for coalition-based approaches.

[9] Some other examples of broad, coalition-based initiatives: The Fair Labor Association monitors labor standards in factories and farm settings; the Roundtable on Sustainable Palm Oil enforces social and environmental standards for palm oil production. Some other examples of targeted, coalition-based initiatives: The Life and Building Safety Initiative enforces building safety in apparel and homegood supply chains; the International Cocoa Initiative aims to reduce child labor in cocoa production in West Africa.

[10] Amengual and Distelhorst (2019) is a recent exception; they use a regression discontinuity design to study Gap Inc's supplier code of conduct for labor. They find that a failing audit grade improves compliance, but only if coupled with the threat of a reduction in Gap's orders.

rely on Bangladesh for its combination of low prices and large production capacity (McKinsey & Company, 2011). Apparel has also driven Bangladesh's industrial transformation (Central Intelligence Agency, 2016); in 2016, apparel exports constituted 81% of its exports and 13% of its Gross Domestic Product.[11] The sector directly employs between 4-5 million of Bangladesh's 66.6 million workers.

As with China, Vietnam, and other leading apparel producers, Bangladesh's apparel sector has been criticized for its poor working conditions and limited freedom of association (FOA) rights for years (ILO, 2016).[12] In a 2011 McKinsey survey of western buyers, buyers list lack of social compliance and economic and political instability as two of the top five risks to sourcing from the country (McKinsey & Company, 2011). Decades of rapid industrial growth and weak state institutions culminated in a series of high fatality industrial accidents in 2012-13, including the collapse of the Rana Plaza building, that killed at least 1,273 workers and injured at least 3,812 workers at exporting factories (Solidarity Center, 2016). [13]

2.3 Private enforcement of safety laws

Prior to the collapse, MNCs sourcing from Bangladesh monitored their suppliers through their own social compliance programs (Bustillo, Wright and Banjo, 2012). In the aftermath, however, MNCs faced pressure to take collective action. European buyers signed an agreement with labor unions, which was known as the Accord on Fire and Building Safety in Bangladesh (hereafter, the Accord). Several U.S. retailers refused to sign the Accord; they formed the Alliance shortly thereafter.[14]

The Alliance was a coalition of 29 multinational retail and apparel firms (e.g., Wal-Mart, Costco, Target, Gap). Its members represented the majority of North American garment imports from Bangladesh. They committed to a five-year agreement to improve the safety performance of their Bangladeshi supplier bases, which included between 600-700 factories and 1.21 million workers. The paper's Supplementary Materials includes

[11] Author's calculations using data from the World Trade Organization and the World Bank.

[12] Garment sector jobs are not without benefits to Bangladeshi society. Heath and Mobarak (2015), for example, show that the growth in these jobs contributed to decreasing fertility, increasing age at marriage, and increasing educational attainment among Bangladeshi girls in recent decades.

[13] 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 western governments penalized the country by removing trade benefits (Greenhouse, 2013*a*). The GoB promised to introduce labor reform and to work with the International Labor Organization (ILO) and other stakeholders to prevent another tragedy.

[14] Their refusal was due to labor unions' participation and to the clause that buyers are subject to legally-binding arbitration (Greenhouse, 2013*b*).

an overview of the Alliance's Member Agreement, its programs, and the nature of this research collaboration.[15] As per its Members' Agreement, the Alliance ceased operations on December 31, 2018. Upon its exit, many members supported the formation of Nirapon, an organization tasked with a similar set of safety oversight functions. 22 out of the 29 Alliance Members joined Nirapon, which continues to operate as of early 2021.

The Alliance required all supplier factories to participate in its building safety audit and remediation and worker training and empowerment programs. Failure to comply with one or more programs resulted in suspension from all Alliance Members' supplier bases; the Alliance suspended 179 factories over its five-and-a-half year term.[16]

2.4 Safety Committees

Safety Committees are a core component of international labor standards for safety. The International Labour Organization (ILO) included SCs in its core Occupational Safety and Health Convention (No. 155, Articles 19-20), which was adopted in 1981.[17] While global statistics on safety committee mandates do not exist, 69 countries have ratified this Convention, and many others mandate SCs (e.g., Bangladesh). The rationale for SCs is to ensure cooperation between managers and workers to achieve a reasonably safe workplace and to ensure that the employer fulfills its obligations (ILO, 1981).

At the time of the Rana Plaza collapse, Bangladesh's labor law lagged behind international standards for safety and for freedom of association (FOA) rights. Many observers directly linked the series of fatal industrial disasters to workers' inability to organize labor unions and to otherwise hold employers' accountable for unsafe conditions (Compa, n.d.). Under intense pressure, in July 2013, the GoB amended the labor law to strengthen safety and FOA provisions.[18] The requirement for SCs was one of the amendment's key provisions.[19] It was strongly resisted by factory owners, who perceived SCs as a step toward collective action by workers. Reflecting owners' resistance, it took the GoB more than two years to publish implementation rules for SCs, even under intense pressure from

[15] More information is also available on the Alliance's website: www.bangladeshworkersafety.org.

[16] The Alliance was also a member of a Private sector-GoB Factory Closure Panel for cases of imminent danger due to structural integrity, which fully or partially closed 35 factories that supplied to the Alliance.

[17] ILO Conventions are legally-binding treaties that may be ratified by member states. Recommendation 164, a non-legally binding guideline, provides detailed guidelines on how SCs should be formed, what types of authority they should have, and what types of legal protection they should enjoy.

[18] The GoB also agreed to a a multi-stakeholder action plan to strengthen its capacity and to improve the sector's safety outcomes (Ministry of Labour and Employment, 2013). To fulfill the action plan, the GoB closely coordinated with the ILO, the Accord, and the Alliance. The Accord and the Alliance were responsible for overseeing safety for the 60-70% of the sector that they covered. The GoB, with the ILO's support, was responsible for the remaining 30-40% of the sector (ILO, 2017).

[19] The mandate for SCs applies to establishments with 50 or more workers.

the international community (Munni, 2015). The implementation rules articulate requirements for SCs' formation, operations, and responsibilities (see Table I).

Table I: Key Safety Committee Requirements

Category	Requirements
Formation	6-12 committee members depending on factory size
	Equal worker-manager representation
	Appointment of worker representatives by collective bargaining agent or Participation Committee*
	SC president appointed by management, SC vice president appointed by worker representatives
	In establishments with >33% female workforce, at least 33% of worker representatives must be female
Operations	Establishments must maintain a written policy on the SC
	SCs must meet at least once per quarter
	SCs must maintain written meeting minutes
	Employers must provide SC members adequate time during working hours to fulfill their duties
	Employers must provide SC members with occupational health and safety training
Responsibilities	SCs must implement factory risk assessment at least once per quarter
	SCs must make safety-improvement recommendations to the employer
	SCs must arrange training and awareness-raising for workers
	SCs will participate in the oversight of the following safety management systems: Management of equipment and work procedure; Management of dangerous fumes, explosives, or flammable items; Fire safety management; Management of dangerous operations, occupational disease, poisonous disease; Emergency Planning
	SCs will investigate accidents and occupational disease and can submit recommendation to employer for treatment and compensation
	SCs will organize regular fire, earthquake, and other disaster management drills

*In factories with a collective bargaining agent (CBA), the CBA selects worker representatives to the safety committee. In factories where there is not a CBA, a Participation Committee (PC) selects worker representatives to the safety committee. A PC is legally required for all factories with 50 or more workers located outside of Export Processing Zones (EPZs). A PC has equal worker-manager representation that aims to promote trust and cooperation between employers and workers. It also aims to ensure application of labor laws.

Source: Translation based on Government of Bangladesh (2015).

Despite the *de jure* requirement to implement SCs, *de facto*, enforcement was low. According to an ILO (2017) report, from 2015-2017, the GoB's focus was primarily on physical safety remediation.[20] Unsurprisingly, compliance was also low (Munni, 2017).

Consistent with the recognition that increased worker voice would be needed to prevent future industrial disasters in Bangladesh, both the Alliance and the Accord Agreements centrally feature worker voice and empowerment initiatives.

[20] As of mid-2017, the ILO had supported the GoB to form SCs at 210 of the 1,549 garment factories under the government's purview (i.e., not including Alliance or Accord-covered factories).

2.5 The Alliance's Safety Committee Program

The Alliance considered credible SCs to be vital to achieving worker empowerment and effective communication between workers and managers (Alliance, n.d.). It created its SC Program to enforce the SC law; the SC Program included five phases:[21]

1. The Alliance verified that factories' worker representation body (Participation Committee (PC), Worker Welfare Association (WWA), or trade union) were democratically selected and formed according to the labor law.[22]
2. Once the PC/WWA/trade union was compliant, the Alliance verified whether the factory had a SC, and if so, whether it was selected and formed according to the law. If necessary, the factory reformed its SC through a compliant process.
3. The Alliance verified that the SC was formed correctly. Once this was confirmed, it provided SC members with training on their roles and responsibilities, on occupational safety and health, and on leadership and communication skills.
4. The SC prepared an action plan for required activities.
5. Once the Alliance approved the action plan, the Alliance intensively monitored the SC on its completion.

In phase 1, the Alliance conducted onsite verification visits to inspect whether the body that is responsible for appointing worker representatives to the SC was legally compliant. Often, these bodies were not democratically/correctly formed, and the Alliance coordinated with the brand(s) sourcing from the factory to oversee a new election. Once the factory reformed its PC, WWA, or trade union, it submitted evidence of its compliance via email and phone to the Alliance. At this point, the Alliance instructed the factory to form or to reform its SC and to submit evidence of its compliance by email (phase 2).

In phase 3, the Alliance again verified that the SC was formed correctly via email and phone. If necessary, it worked with factories to reform the SC. It then provided two days of training to two worker and two management representatives from the SC.[23] During training, the Alliance set expectations for the SC Program's central feature: The preparation and fulfillment of an action plan to achieve compliance with the SC law (phases 4-5).

The action plan used an Alliance template and included a detailed schedule of required activities. Several members of management, the SC President, and the SC Vice President had to sign off on it. Before approving the plan, the Alliance reviewed it and

[21] Nirapon continues to require supplier factories to participate in a similar SC Program.

[22] PCs are legally-required for all factories with 50 or more workers outside of Export Processing Zones (EPZs). In EPZs, the worker representation structure is a Workers' Welfare Association (WWA). In most Alliance supplier factories, the PC is the body that appoints the worker representatives to the SC.

[23] SCs range from 6-12 members, with even worker-manager representation, depending on factory size.

worked with the factory to make revisions. The factory then implemented the plan and provided evidence to the Alliance by e-mail within 2-3 days of each required activity. Repeated failure to submit evidence resulted in escalation of the factory's status toward suspension. The Alliance reviewed submissions and investigated by phone calls, e-mails, and onsite audits that were unannounced or announced within a certain time period. At the end of the six-month period, the Alliance reviewed the factory's progress. If the Alliance found it to be insufficient, the factory could be required to repeat parts of the program or its status could be escalated toward suspension. If found to be sufficient, the factory returned to the pool of factories being monitored through the Alliance's general monitoring program.[24]

In the RCT, I experimentally manipulate factories' exposure to phases 3-5 of the SC Program. This is because the Alliance's authority did not extend to the PC/WWA/trade union, so phases 1-2 depended on Alliance brands' engagement with their suppliers. Organizing free, fair, and contested democratic elections for worker representation bodies often took many months. As such, factories became eligible for the RCT once this process was complete. The treatment effects that I identify are thus the effects of enforcing the SC regulation on factories with a SC that exists, at least on paper, at baseline.

3 Conceptual framework

In this section, I clarify how MNCs' enforcement objectives for their suppliers' compliance may diverge from a traditional state regulator's. I then present a simple framework for a factory manager's compliance decision under MNC enforcement. The fact that MNCs' true motivation may be for the manager to achieve formal compliance, rather than meaningful safety improvements, creates the possibility for the manager to engage in box ticking: Cheap compliance that does not improve safety. For example, the SC law requires factories to maintain a list of SC members. As it's cheap to select a committee and to record a list of names, but does not obviously improve safety, I consider this a cheap form of compliance. The law also requires that SCs conduct quarterly risk assessments and submit recommendations on resolving hazards to management. As it's costlier to conduct risk assessment, but arguably more effective at improving safety, I consider this a costly form of compliance. I generate different testable predictions for a manager's choice of cheap and of costly but effective compliance depending on MNCs' willingness to penalize noncompliance and to exert monitoring effort.

[24]More information about the SC Program, including many of the Program's materials, is available on the Alliance website.

I also suggest that the potential for complementarity between state enforcement capacity and managerial capacity of the private sector could provide insights into low levels of compliance in developing countries. In the model, I introduce heterogeneous managerial capacity to see how it affects the manager's decision.[25] The model generates testable predictions about factories' baseline compliance, the effects of MNC enforcement, and how these effects may depend on managerial capacity.

3.1 MNC versus state-supplied enforcement

In models of compliance decisions, enforcement regimes are typically characterized by the probability of noncompliance being detected and the penalty for noncompliance (Becker, 1968). The regulator is assumed to be the state. In contrast, in my framework, MNCs determine monitoring and penalties, and they may aim for suppliers to achieve formal compliance, rather than to meaningfully improve safety. If compliance is costly for suppliers, MNCs may prefer to keep monitoring and/or penalties low. If monitoring is costly, MNCs may prefer to provide a low level. If punishing or changing suppliers is costly, even if MNCs observe low compliance, they may prefer not to penalize it.

The more observable aspect of MNCs' enforcement regimes is monitoring, which as discussed in Section 2, is typically non-zero. The extent to which MNCs penalize suppliers for noncompliance, however, is much more obscure. As such, I generate testable predictions for the manager's choice of compliance depending on the magnitude of penalties imposed by MNCs relative to the manager's costs of increasing compliance.

3.2 The manager's compliance decision

The manager chooses his extent of costly, effective compliance $k_1 \in [0,1]$ and cheap, box ticking compliance, $k_2 \in [0,1]$. $k_i = 0$ means full noncompliance and $k_i = 1$ means full compliance, for $i \in \{1,2\}$. I assume that the costs of compliance are increasing and convex; for simplicity, I assume that they take the form: $\frac{c_i k_i^2}{2}$. Effective compliance is costlier than box ticking: $c_1 > c_2 > 0$. Only costly compliance achieves true compliance with the law and improves safety; factories use cheap compliance to help achieve formal compliance, but it does not improve safety. In my simplified framework, the benefits of effective compliance accrue to workers and/or to MNCs; they are not internalized by the manager.

I also assume that it is cheap for MNCs to observe box ticking by their suppliers: Suppliers can submit documentation at very low cost. As such, I assume that MNCs effort-

[25] I note that in the RCT, managerial capability is not an aspect of the randomization; I test for heterogeneous treatment effects by suppliers' baseline managerial practices.

lessly, perfectly observe k_2. Their motivation to monitor less observable but more effective k_1 is parameterized by $\theta \in [0,1)$. Finally, I assume that MNCs penalize noncompliance with penalty $p > 0$. MNCs penalize both forms of noncompliance at the same rate, for example, to protect their reputation in case of a bad media report or an industrial disaster: While MNCs may not truly value safety, they can claim that they penalize suppliers who appear noncompliant.

The manager chooses the optimal levels of k_1 and k_2 in order to maximize the value of the following equation:

$$\max_{k_1,k_2} V(k_1,k_2) = -\frac{c_1 k_1^2}{2} - \frac{c_2 k_2^2}{2} - p(1-k_2) - \theta\, p(1-k_1) \tag{1}$$

Proposition 1. For given costs of costly and cheap compliance (c_1 and c_2), penalty rate (p), monitoring effort (θ), the optimal levels of costly and cheap compliance (k_1^*,k_2^*) are:
Case 1: $p < c_2 < c_1$:

1. If $\theta = 0$: $k_1^* = 0$. $k_2^* = \frac{p}{c_2}$.
2. If $\theta \in (0,1)$: $k_1^* = \frac{\theta\, p}{c_1}$ and $k_2^* = \frac{p}{c_2}$.

Case 2: $c_2 \leq p \leq c_1$:

1. If $\theta = 0$: $k_1^* = 0$. $k_2^* = 1$.
2. If $\theta \in (0,1)$: $k_1^* = \frac{\theta\, p}{c_1}$ and $k_2^* = 1$.

Case 3: $c_2 < c_1 < p$:

1. If $\theta = 0$: $k_1^* = 0$. $k_2^* = 1$.
2. If $\theta \in (0, \frac{c_1}{p})$: $k_1^* = \frac{\theta\, p}{c_1}$ and $k_2^* = 1$.
3. If $\theta \geq \frac{c_1}{p}$: $k_1^* = k_2^* = 1$.

See Online Appendix (hereafter, OA) A for the proof. The size of the penalty relative to the costs of cheap and costly compliance partially determines the manager's optimal levels of cheap and costly compliance. When the penalty is relatively low ($p < c_2$), there is an interior solution for cheap compliance, and costly compliance is zero or interior depending on MNCs' monitoring effort. When the penalty is relatively higher ($c_2 \leq p \leq c_1$), the manager fully complies with cheap requirements, and his engagement in costly compliance is zero or interior depending on MNCs' monitoring effort. When the penalty is high ($c_1 < p$), the manager fully complies with cheap requirements, and his engagement

in costly compliance is interior or full depending on MNCs' monitoring effort. Unless the manager fully complies with both cheap and costly requirements, his engagement in cheap compliance is strictly greater than that of costly compliance.

When MNCs increase monitoring (increase in θ), as long as $k_1^* < 1$ at baseline, it will increase. The manager will only achieve full costly compliance, though, if both penalties and monitoring are high ($p > c_1$ and $\theta \geq \frac{c_1}{p}$). k_2^* will not change, so the gap between cheap and costly compliance levels will partially or fully close.

Higher ability managers may have a lower marginal cost of compliance. For example, if $\lambda > 1$ is a manager's type, costs may take the form $\frac{c_i}{\lambda}$, for $i \in \{1, 2\}$, and higher ability managers have higher values of λ. In this case, if MNCs increase monitoring, higher ability managers' optimal level of costly compliance will increase more than lower capability managers'. There will be no change in cheap compliance for either type. See Appendix 6.

3.3 Additional predictions

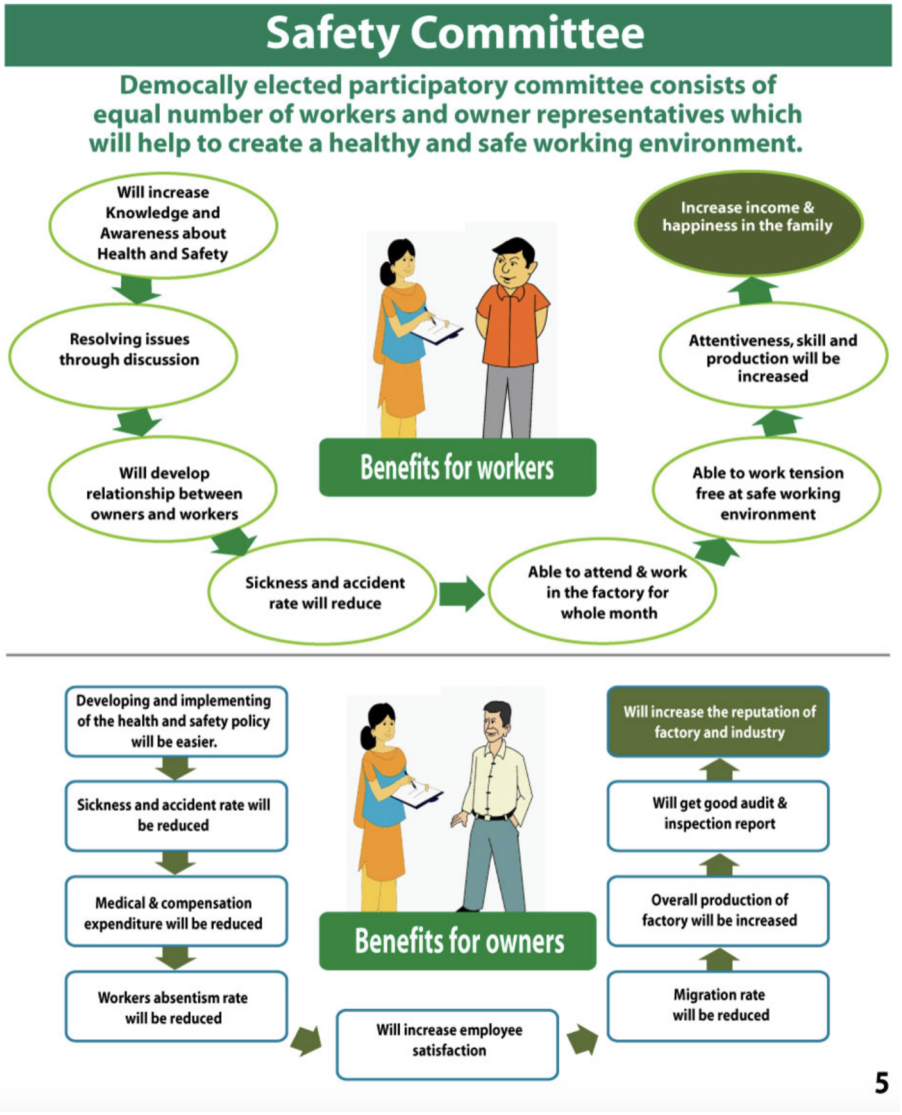
If MNC enforcement increases costly compliance, a natural next question is how this affects workers' safety and well-being. Figure I displays the Alliance's proposed causal chain. The chain suggests that SCs affect these outcomes through increased knowledge and awareness, through improved relations with management, and through improved factory safety. To test these possibilities, I first test objective measures of workers' safety, including physical indicators of factory safety, workers' safety knowledge, and workers' visits to their factories' medical clinics. I then turn to subjective measures of workers' well-being. These measures may be positively affected by safety improvements, but this does not directly follow if, for example, more stringent safety protocols impose costs on workers or if employers reduce other benefits in response.

Next, I turn to the economic impacts on supplier factories. First, I examine effects on labor productivity. As displayed in Figure I, there are multiple mechanisms through which implementation of SCs may improve labor productivity. For example, workers may make fewer visits to the medical clinic or miss work less. It's unclear, though, whether SCs' productivity-enhancing effects outweigh their productivity-decreasing ones. As with broader OSH regulations, there is a concern that SCs reduce production speed and/or capacity, for example, due to stringent safety protocols and floor-plan requirements. Further, barring increases in working hours, SCs' meetings and activities with workers reduce the amount of time allocated toward production. In sum, it is ambiguous whether, on net, SCs positively or negatively affect labor productivity.

Second, I examine the effects on wages and employment. These depend on SCs' net

costs and benefits to factories and on the extent to which labor markets are competitive.

Figure I: Alliance SC Theory of Change



Source: Alliance training materials for SC members (English translation).

A primary concern is that well-intentioned enforcement of SCs may actually make workers worse off overall by adversely affecting wages and employment. Suppose that on net, SCs increase factories' costs and labor markets are competitive. In this case, employers lower wages, and if SCs' costs exceed workers' valuation of them, employment will even-

tually fall.[26,27] If employers have some power in the labor market, however, barring SCs' imposing extremely high costs to employers, wages and employment will not fall, and under less likely conditions, may actually increase (Manning, 2003).

4 Research design

4.1 Randomized assignment to the SC Program

The RCT was built into the roll out of the SC Program.[28,29] From January-December 2017, when the Alliance had a batch factories with verified PCs/WWAs/trade unions, it sent me the list. Within batch, I randomly assigned 50% of factories to the treatment group and 50% to the control group. The result is a stratified RCT with six strata (batches), and a total of 41 treatment factories and 43 control factories (84 total factories).[30] In 11 cases in which multiple factories shared ownership and location (building or compound), I randomly selected one factory to participate in the RCT.[31] If the factory was assigned to the control condition, the Alliance did not conduct the SC Program with any other factories at the same location. OA Table CI reports summary statistics.

4.2 Data collection and measurement

This analysis uses three main sources of data. First, it uses several types of data collected during three separate, day-long visits to factories implemented over nearly one year. Second, it uses monthly production, human resource, and other business performance-related data collected using a retrospective questionnaire administered following the final data collection visit. Third, it uses administrative data from the Alliance. The Alliance invited factories to cooperate with data collection.

The onsite visits included three types of data collection: Surveys of stakeholders, document collection and verification, and spotchecks of safety conditions. Surveys included

[26]This study was implemented from January 2017-December 2018. In relation to the study period, Bangladesh's minimum wage was increased to 5300 Bangladeshi Taka per month (US$63) in 2013. It was not increased again until December 2018, which was the final month of data collection for 11 study factories.

[27]Wages may also directly fall if compensation includes production-based incentives. While it varies across factories, compensation often includes a base wage and some degree of production-based incentives. If the intervention lowers productivity, wages could be directly negatively impacted.

[28]The Alliance rolls out all of its programs in phases, so from the experimental factories' perspective, it would not be apparent that the factory was part of a treatment or control group.

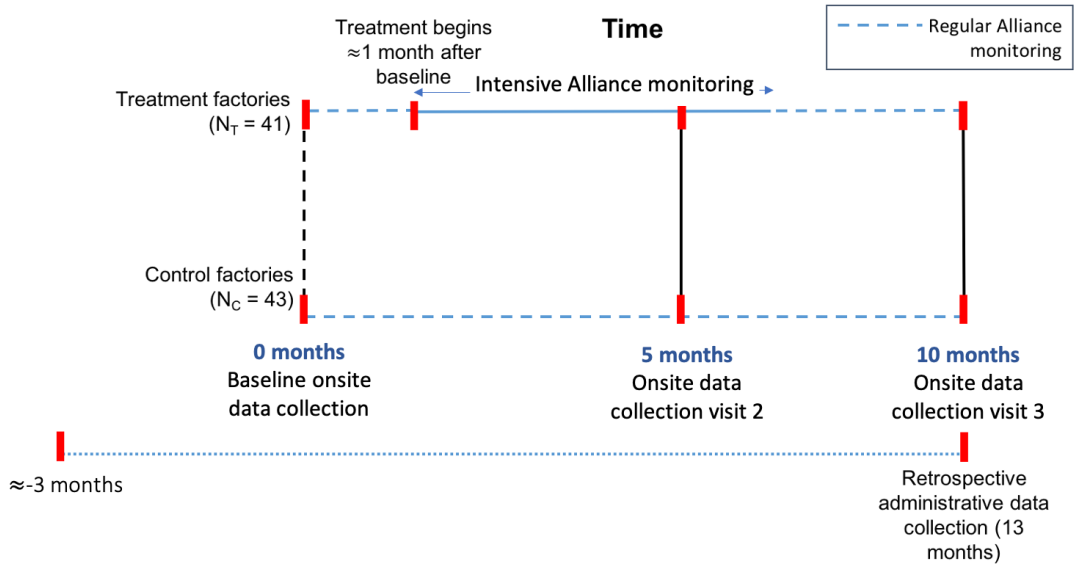
[29]The Alliance implemented the SC Program with factories that did not supply to Accord buyers because the Accord also implemented a SC Program with its suppliers, including those covered by the Alliance.

[30]All control factories were required to participate in the SC Program after completing the study period.

[31]A compound is a plot of land housing multiple factories at the same address.

20 randomly selected workers, the SC President, two randomly-selected SC worker representatives, the factory's most senior manager, and up to 20 randomly-selected lower-level managers. The document verification process entailed checking legally-required and Alliance-required documentation. It also included photographing records for digitization by the research team. At the second and third visits, a trained assessor visited the production floor to check physical safety conditions using a checklist. The team leader was an assessor, who was responsible for managing interactions with management, verifying documentation, and implementing the safety spotchecks. A junior assessor oversaw the the survey process, photographed records, and supported survey implementation. Three enumerators conducted surveys.

Figure II: RCT timeline



Source: Notes: In the figure, the solid blue line for treatment factories denotes the SC Program period (intensive monitoring phase). The dashed line for both treatment and control groups denotes regular Alliance monitoring, under which factories receive indirect monitoring through other Alliance programs (e.g., building safety remediation verification visits) and those factories that have completed the SC Program may be audited through onsite visits. Due to the staggered entry of strata of factories into the RCT, the calendar timeline varies by stratum. The earliest baseline data collection visits occurred in January 2017.

In any study of compliance, one must be concerned about subjects' incentives to misreport. I designed the data collection protocols to minimize experimenter demand effects and the potential for the SC Program to affect reporting. I also collected data that would allow me to directly test for truth-telling. In OA B, I provide econometric evidence against reporting bias affecting my estimated treatment effects. In the Supplementary Materials,

I provide a detailed overview of how the data collection protocol was designed to minimize experimenter demand effects.

Figure II displays the experiment's timeline. The first visit established factories' baselines. The second visit, approximately five months later, aimed to measure outcomes immediately after treatment factories completed the most intensive portion of the SC Program. The third visit, approximately 10 months after baseline, aimed to measure outcomes a few months after treatment factories completed the SC Program.

4.2.1 Pre-specified primary outcomes

I analyze two groups of primary outcome variables. The first measures the intervention's effects on compliance and safety outcomes, while the second measures its economic effects. Beginning with the first group, the outcomes are: (1) Compliance with Bangladesh SC Regulation (index); and (2) safety indicators (index). A complete list of all index variables' sub-variables is available in the paper's Supplementary Materials.

The first outcome is a standardized index variable that summarizes factories' compliance with the SC regulation. I use an index variable because compliance with the SC regulation is many-dimensional. The regulation includes three categories of requirements, including how SCs are formed, how they operate, and their responsibilities (see Table I). To determine the variables included in the index, I enumerated the regulation's stipulations. Whenever relevant, I measure a factory's compliance with a stipulation using multiple sources information. For example, to determine how worker representatives to the SC were selected, I put 50% weight on the SC president's report (a member of management) and 50% weight on the mean of SC worker representatives' reports.[32]

Consistent with the model, I also divide the compliance index into cheap and costly sub-indexes. I aim to include requirements that may be considered box ticking and/or are easily falsified in the cheap compliance index; for example, maintaining a list of SC members (sub-variable no. 6). I aim to include requirements that are clearly costlier and/or more difficult to falsify in the costly compliance index; for example, the requirement that the SC conduct regular risk assessments of the factory (sub-variable no. 20). Sub-variables' categorization as cheap or costly is available in the paper's Supplementary Materials. This analysis was not pre-specified and should be interpreted as exploratory.

The second outcome measures safety indicators that capture SCs' effectiveness at fulfilling the law's intent. The regulation prescribes responsibilities for SCs related to man-

[32] I use this weighting scheme for all index variables with reports from the SC president and from the SC representatives. At baseline, the correlation between presidents' and workers' reports ranges from -0.16 (sub-variable no. 23) to 0.38 (sub-variable no. 3); most correlations are positive.

agement of physical factory safety, to training workers, and to safety culture. Correspondingly, the index includes both physical and cultural indicators of safety:

- Physical safety:
  - Performance on an independent spotcheck of factory safety conditions.
  - Progress with required building safety remediation based on Alliance building safety audits (Alliance "Corrective Action Plan (CAP)" completion).[33]
- Factory safety culture:
  - Workers' awareness of the SC.
  - Workers' safety knowledge.
  - Senior managers' awareness of the SC.

In light of the wide-ranging safety oversight that the SC regulation assigns to the SC, I worked with an OSH expert to determine the spotcheck checklist. The expert helped me to identify critical items from a checklist for typical OSH audits of the factory floor that a trained social compliance assessor could check during a 30-minute floor visit.[34]

The ideal measure of SCs' effects on OSH would be injuries and illnesses. There are multiple reasons, though, why I do not use these as my primary outcome. First, the intervention aimed to empower workers to raise safety issues and concerns. Consequently, on net, it could increase reported injuries and illnesses, even if it reduced their true rates. While this problem affects all sources of information about injuries and illnesses, I identify factories' medical clinic visitor records as the source that is most likely closest to the truth. I provide supporting evidence from analysis of these records.[35]

Turning to the economic primary outcomes, these include: (3) Workers' job satisfaction

[33]Every Alliance-audited factory had a Corrective Action Plan (CAP) based on violations found in the Alliance's building safety audit. The CAP detailed the remediation actions that the factory would take to address the safety violations. The Alliance monitored factories' remediation progress and suspended factories that failed to make sufficient progress.

[34]We excluded items that the SC could not plausibly influence within the study's duration. We also identified several items that the social compliance assessor would only check during the 9-10 month visit. The rationale for this approach was twofold: First, the OSH expert identified eight items that required more than 3-4 months, but plausibly less than 8-9 months, for the SC to address. Second, I wanted to reserve some factory spaces (e.g., bathrooms) to only be visited during the third visit. I aimed to gain insight into the extent to which management was responding to the research team's visits. Due to an administrative error, the eight items were not included in the third visit checklist for 14 out of 80 factories. As such, I depart from my PAP by not including these items.

[35]I am only able to analyze records for 62 factories. This is because numerous factories, in particular those located in EPZs and in multi-factory compounds, use centralized facilities and do not maintain factory-specific records. Among factories that do maintain records, their information content and legibility varies.

and mental well-being (index); (4) labor productivity;[36] (5) employment; and (6) wages. Beginning with outcome no. 3, it is an index variable that summarizes the enforcement interventions' effects on self-reported and revealed preference measures of workers' job satisfaction and mental well-being. I construct it using survey questions and administrative data on worker turnover and absenteeism.

The final three primary outcomes measure the intervention's effects on factories' business competitiveness. Labor productivity is measured as the log of the physical quantity of output (e.g, pieces of clothing) per person-hour. Person-hours are calculated as number of workers times the average weekly working hours times 4 weeks per month plus the number of management-level employees times average weekly working hours for managers times 4 weeks per month.[37] In six factories that produce multiple products, output is measured at the product-level. For these factories, I include the main product in the analysis and determine the share of labor allocated to this product using employee lists.[38] Employment is the total number of people employed at the factory in a month. Finally, wages are the log of gross wages paid to all employees in a month. These three outcomes are measured using administrative data provided by the factories.

To construct the index variables, as per my PAP, I use the methodology proposed by Anderson (2008) based on O'Brien (1984).[39] The method entails an average of a family of variables that have each been oriented to be unidirectional, standardized, and weighted by the sum of its row in the inverse variance-covariance matrix calculated using the control group. The weighting maximizes the amount of information captured by the index, as it places less weight on highly correlated outcomes and more on less correlated ones. This approach is well-suited for this study because, due to the staggered roll-out, I was not able to collect a full baseline before committing to the construction of indexes. Summary index variables also reduce the number of hypothesis tests, which reduces the risk of overrejection of the null hypothesis. Finally, they increase my ability to detect effects on multiple outcomes that, aggregated, achieve statistical significance (Anderson, 2008).

I also pre-specified secondary outcome variables to explore possible mechanisms underlying the effects on primary outcome variables. Index components for secondary in-

[36] I pre-specified that I would analyze total factor productivity or labor productivity. I indicated that I would analyze labor productivity if I determined that I could not measure non-labor inputs to production with sufficiently high quality. Ultimately, I decided that I could measure labor productivity for more factories and with less measurement error.

[37] OA Table CIII presents treatment effects on the log of the physical quantity of output and average weekly working hours.

[38] I determine a factory's primary product using quantities of physical output.

[39] Other recent RCTs that report index results using this methodology include Casey, Glennerster and Miguel (2012) and Haushofer and Shapiro (2016).

dex variables are available in the paper's Supplementary Materials, OA Table CII reports baseline balance tests, and OA Table CIII present results. I reference these results to support interpretation. In the interest of transparency, I report all deviations from my PAP and their rationales in OA Table CIV. Overall, I adhere very closely to my PAP.

4.2.2 Econometric analysis

I estimate the intervention's average treatment effects using the following specification:

$$Y_j = \alpha + \beta\, T_j + \theta\, Y_{j,t=0} + \gamma_w + \mathbf{X}_j'\lambda + \epsilon_j \tag{2}$$

where Y_j is the outcome for factory j. T_j is the treatment indicator, $Y_{j,t=0}$ is a control for the baseline value of the outcome variable. γ_w is a stratum indicator, and ϵ_j is the residual. β is the coefficient of interest. I report robust standard errors; for outcomes with multiple observations per factory, I report standard errors that are clustered at the factory level. Given my small sample size, I also report randomization inference (RI) p-values for all primary outcomes (Athey and Imbens, 2016, Heß, 2017). I also estimate effects for primary outcomes using the post double selection (PDS) lasso to select control variables (Belloni, Chernozhukov and Hansen, 2014).[40] The set of potential controls include all variables in Table II Panels B and C that are available for the full sample, the log baseline number of employees, and the baseline value of the outcome variable.

To test for heterogeneous treatment effects, I use the following regression model:

$$Y_j = \alpha + \beta_1\, T_j + \beta_2\, R_j + \beta_3\, T_j * R_j + \theta\, Y_{j,t=0} + \gamma_w + \epsilon_j \tag{3}$$

where R_j is an indicator for above median baseline value of a pre-specified interaction variable. The notation for equation 3 is otherwise analogous to that for equation 2. In this specification, β_1 is the estimated treatment effect on factories with a below median baseline value of the interaction variable, $\beta_1 + \beta_3$ is the estimated treatment effect on factories with an above median baseline value of the interaction variable, and β_3 is the difference between these two effects. I report β_1 and $\beta_1 + \beta_3$ as well as the p-value for β_3.

To account for multiple hypothesis testing, I report multiplicity-adjusted p-values. Across my primary outcome variables, I control the False Discovery Rate (FDR), the expected proportion of rejections that are false positives. I report FDR-sharpened p-values for my preferred specification for all primary outcomes (Anderson, 2008). For index vari-

[40] This approach has two advantages: It allows me to increase the precision of my estimates while avoiding concerns about specification searching, and it allows me to test my results' robustness to the possibility that chance baseline imbalances between the treatment and control group influence my estimates.

ables, I also show p-values adjusted to control the FDR across each variable's sub-indexes.

4.2.3 Internal Validity

Baseline Balance: Table II shows balance tests for primary outcomes and factory and worker characteristics. Certain variables are not available for all factories. In particular, factories that attrited from the sample did not provide data on business outcomes. Among non-attrited factories, five declined to provide production data, and eight declined to provide wage data. In sum, the randomization successfully generated two groups that are balanced along observable characteristics. There is one variable with a statistically significant difference at the 10% level, which is the proportion of surveyed female workers. This difference is not statistically significant among non-attrited factories.

Although the difference is not statistically significant, treatment factories' performance on the job satisfaction and well-being index is 0.11 sds lower than controls'. This gap is largely due to a treatment factory whose performance negatively deviates markedly from other factories' - its index value is more than 4 sds below the mean. I adopt a common approach to handling outliers, which is to present results including the outlier and to show baseline balance and the main the results after dropping the outlier (OA Tables CV and CVI). The results are robust to dropping this factory.

Finally, turning to labor productivity, although not statistically significant, there is a qualitatively large difference between treatment and control factories. This difference is due to small differences in factory types between the groups. The treatment group has somewhat more non-sewing factories (e.g., washing and accessories) that tend to be more capital intensive. For this reason, I also show that there are no differences in labor productivity between treatment and control factories that produce the same type of product.

OA Table CVII shows balance tests for SC president, SC worker representative, and senior manager survey participants; variables from these surveys are used to construct certain index variables. The treatment and control groups are balanced on all variables.

Compliance: Three treatment factories did not receive treatment by the second data collection visit. Once we identified the reasons for the delays, we resolved them for other factories that could have been impacted.[41] A fourth factory began the SC Program less than two weeks before its second data collection visit. I address non-compliance by presenting Intent to treat (ITT) estimates and presenting the Local Average Treatment Effect (LATE) estimates for primary outcomes in OA Table CVIII.

[41] Two factories are located in the Chittagong Region, where the Alliance implemented the SC Program in batches to ensure cost effectiveness, and it did not have a sufficient number of factories to implement it with these factories. One factory did not participate due to a critical manager being on an extended leave.

Table II: Baseline balance tests

	Control mean (SD)	T-C diff	p-value [RI p]	Number of factories
	(1)	(2)	(3)	(4)
Panel A: Primary outcome variables				
Compliance index	0.000	-0.072	0.448	84
	(0.276)		[0.453]	
Cheap SC compliance	0.000	-0.106	0.403	84
	(0.414)		[0.418]	
Costly SC compliance	0.000	-0.029	0.744	84
	(0.355)		[0.732]	
Safety indicators index	0.002	-0.053	0.627	84
	(0.401)		[0.621]	
Job satisfaction & well-being index	-0.005	-0 .113	0.263	84
	(0.369)		[0.257]	
Number of employees	1166	-150	0.593	84
	(1193)		[0.609]	
Gross wages (log)	15.820	-0.190	0.467	72
	(1.053)		[0.471]	
Labor productivity (log)†	1.569	0.399	0.279	75
	(1.424)		[0.254]	
Labor productivity (log)†, product FE	0.019	-0.043	0.783	75
	(0.609)		[0.782]	
Panel B: Factory characteristics				
Trade union at factory (1=Yes)	0.047	-0.045	0.164	84
	(0.213)		[0.514]	
EPZ(1=Yes)	0.163	0.036	0.672	84
	(0.374)		[0.775]	
Sewing (only)	0.465	-0.129	0.242	84
	(0.505)		[0.271]	
Number product types	1.163	0.053	0.641	84
	(0.652)		[0.688]	
Monthly absenteeism (%)	4.859	-0.667	0.439	80
	(4.581)		[0.439]	
Monthly turnover (%)	3.920	0.012	0.991	84
	(4.894)		[0.989]	
Prop. employees visit medical clinic (daily)	0.012	0.005	0.402	62
	(0.0055)		[0.447]	
Participation in Alliance training	0.070	-0.021	0.670	84
(6 mo pre-baseline)	(0.258)		[1.000]	
Number Alliance remediation visit to factory	0.186	-0.014	0.882	84
(6 mo pre-baseline)	(0.450)		[1.000]	
Panel C: Worker survey respondent characteristics				
Age	27.179	0.185	0.815	84
	(3.606)		[0.821]	
Proportion female	0.568	-0.111*	0.082	84
	(0.279)		[0.080]	
Education (yrs)	6.222	-0.428	0.243	84
	(1.585)		[0.233]	
Tenure (yrs)	3.852	-0.180	0.712	84
	(2.412)		[0.707]	
Prior industry experience (yrs)	1.535	0.041	0.860	84
	(0.866)		[0.857]	

Notes: This table reports OLS estimates of baseline differences between control and treatment groups. For each outcome or covariate, I report the baseline control group mean and SD in column (1). In column (2), I report the estimated coefficient for the treatment indicator from a regression of the outcome or covariate on the treatment indicator and stratification variables. In column (3), I report the p-value for the treatment indicator calculated using robust standard errors. I also report the RI p-value for the coefficient reported in column (2) based on 5000 draws. In column (4), I report the sample size for the regression. †The regression sample is trimmed at the 1st and 99th percentile of all factory-month labor productivity observations. *p<0.1; **p<0.05; ***p<0.01.

Attrition: Four factories, two treatment and two control, attrited from the sample. Three of the four were suspended by the Alliance for failure to make progress with physical building safety remediation. One control factory refused to participate in the second data collection visit. I address attrition by reporting Lee (2009) bounds on the treatment effects for primary variables with statistically significant treatment effects (OA Table CIX). There is minimal difference between the upper and lower bounds of the treatment effects, and with the exception of the lower bound for the safety indicators index, all estimates are statistically significant at the 5% or 10% level.

5 Results

I present the intervention's effects in five sub-sections. First, I present the effects of multinational enforcement on factories' compliance and on safety indicators. Next, I assess its effects on workers' job satisfaction and well-being and on factories' competitiveness. Third, I discuss whether the treatment effects are heterogeneous across factories. Fourth, I test for persistence after MNCs cease intensive enforcement. Finally, I discuss the extent to which this study's results generalize to other settings.

5.1 MNC enforcement, factories' compliance, and safety

5.1.1 Factories' compliance at baseline

In phase 1 of the SC Program, factories submitted documentation to the Alliance that they had a free, fair, and contested election for the worker representation body responsible for appointing worker representatives to the SC (see Section 2.5). Once the Alliance received this evidence, it visited the factory to verify that the worker body was compliant. Based on the Alliance's records, I can verify that at least 83 out of 84 factories received this visit.

During these visits, the Alliance also checked the status of the factory's SC. According to its records, 52% of sample factories had an improperly formed SC (38 factories) or no SC (6 factories).[42] In cases of improper SC formation, the SC was often selected by a non-democratic PC/WWA, was mostly managers, or was fully overlapping in membership with the PC/WWA. Among the 40 factories with a correctly-formed SC, I observe that the Alliance interviewed one or more SC worker representatives at 34 factories. In 12 of these (35%), it found that they were unaware or little aware of their responsibilities.

Following this visit, if needed, the factory submitted evidence of (re)forming its SC to

[42] 98% of these visits occurred after the legal deadline for SC formation. The median visit date was 10 months after the deadline.

the Alliance via email (phase 2). At this point, the factory became eligible for this study. The research team's baseline visit took place about 3.25 months later (median factory).

First, I examine baseline compliance, as measured by the research team, using the framework Section from 3. Figure III plots the raw data for factories' cheap and costly compliance. In the Figure, to facilitate interpretation, I show cheap and costly compliance as a proportion between 0 (complete non-compliance) and 1 (full compliance).[43] The Figure shows that sample factories comply with about 80% of cheap requirements, but only 57% of costly requirements, at baseline (p-val of difference: 0.008). The fact that factories engage in costly compliance at baseline is consistent with the Alliance providing some monitoring ($\theta > 0$). The fact that baseline cheap compliance is interior and is closer to full is consistent with the Alliance providing penalties that are material but not too high relative to the cost of cheap compliance ($p < c_2$). It also implies that increased monitoring through the SC Program should not lead to full compliance with costly requirements.

In terms of factories' compliance with the specific requirements for SCs' formation, operations, and responsibilities. All except one factory had a SC at baseline, at least on paper. 20% of factories had established their SC by the legally-required deadline.[44] 73% of SCs were of the correct size and composition, although many SC presidents and worker representatives reported non-compliant selection procedures for worker representatives.[45,46]

In most factories, SCs were just becoming active. In 10% of factories, the SC had not yet met; in a further 16%, the SC had only met once. There were some reports of interference: by management.[47] Many SCs were not implementing their legally-required responsibilities. A key responsibility in the labor law is risk assessment, including making recommendations to resolve issues to management. At baseline, 15% of SCs had ever conducted a risk assessment. 73% of senior managers report receiving reports on safety issues at least least once per three months. SCs' reported fulfillment of other legally-

[43] While I use a standardized index variable for my main analysis, for most variables in the index, I can characterize compliance as an indicator variable or as a proportion. I drop the sub-variable that I cannot represent in this way, the frequency of SC meetings per three months, from the analysis.

[44] Relative to its baseline data collection visit, the median factory established its SC about 5 months prior.

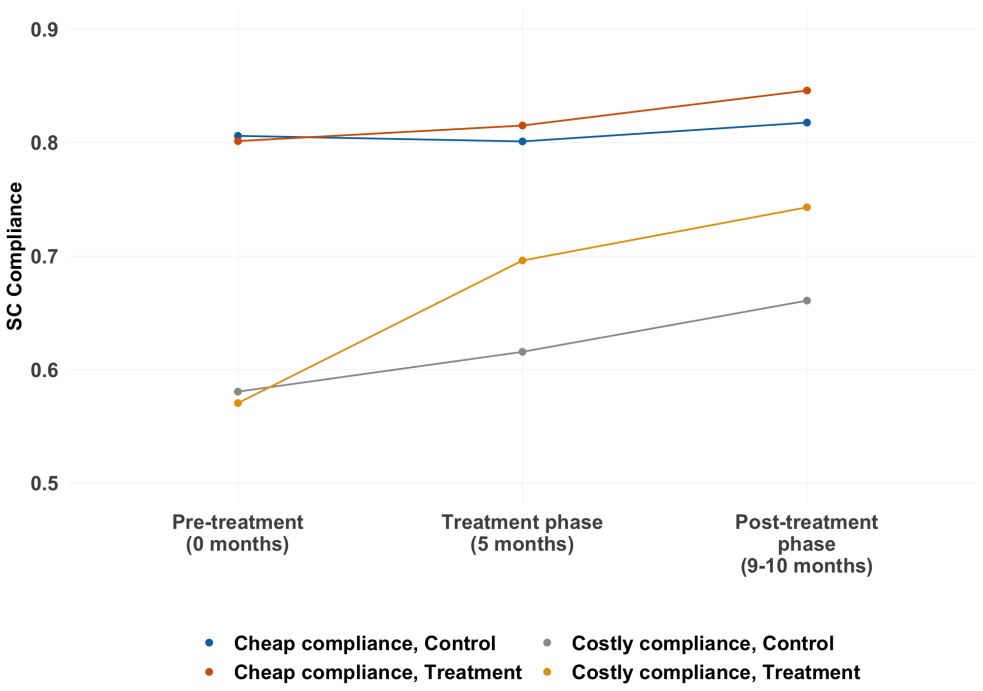
[45] In one control factory, the SC was found to be comprised only of managers. In this case, compliance index outcomes related to correct formation of the committee are coded as non-compliant. At the second visit, the same factory provided the names of workers whom it indicated were members of the SC. Through the SC worker representative survey, it emerged that these workers were not members of the SC. Management had instructed them to participate because the composition of SC remained all managers. Again, the compliance index outcomes related to correct formation of the committee are coded as non-compliant.

[46] Despite these issues, there was high consistency between factory documentation and SC presidents' reports of SC size and composition ($\rho = 0.93$).

[47] In 10% of factories, at least one worker representative reported that management had offered bribes or attempted to block SC activities. 5% of presidents and 7% of worker representatives reported that they were not considered on duty for SC activities.

required responsibilities varied.[48]

Figure III: Cheap & costly compliance (raw data)



Does the SC Program actually increase monitoring?

My conceptual framework assumes that the Alliance's SC Program entails increased monitoring of costly compliance with the SC regulation. Is there empirical support for this assertion? I can partially answer this question using the Alliance's administrative records. Onsite audits reveal more information than remote monitoring but are more costly to conduct. The Alliance conducted audits during and after factories' participation in the SC Program. According to its administrative records, during the audit, an Alliance staff member reviewed SC-related documentation, interviewed SC members and workers, and inspected the factory. Among the treatment factories, 10% were audited during the SC Program and 15% were audited in the six months after completing it. Evidently, the SC Program entailed a material threat of being monitored through onsite audits.

[48] According to SC presidents, SCs were most likely to participate in fire prevention and preparedness activities (84%) and least likely to participate in accident investigation (55%). Although 44 SC presidents reported that the SC was responsible to investigate in case of an accident, only 7 indicated that the SC had actually participated in an accident investigation.

5.1.2 The effects of MNC enforcement on compliance

Table III presents the results for the index of compliance with the SC regulation. Panel A shows that the treatment effect on compliance is about 0.20 sds; the effect is highly statistically significant when estimated with a control for baseline compliance (column (1)) and with PDS lasso selected controls (column (2)). In the latter case, the FDR p= 0.001.

The model predicts that increased monitoring by the Alliance should only affect factories' compliance with costly requirements; cheap compliance is already observable to the Alliance and does not depend on its effort to monitor. Table III columns (3)-(4) show results that are consistent with this prediction: The effect on cheap compliance is small and not statistically significant, while the effect on costly compliance is positive, large, and highly statistically significant. As can be seen in Figure III, consistent with the model's prediction, these effects partially close the gap between factories' compliance with cheap and costly requirements.[49] The fact that it does not close completely, though, is consistent with the Alliance's penalties, while being greater than zero, being lower than the cost of cheap compliance. Together, these results demonstrate that MNC enforcement can cause factories to go beyond box ticking and to increase effective compliance with the law, above and beyond the effects of state-supplied enforcement. They also demonstrate, however, that at least in the context that I study, MNCs may opt *not* to set penalties high enough so as to induce suppliers to achieve full compliance.[50]

Panel A of Table IV displays the results for the formation, operations, and responsibilities sub-indexes.[51] While treatment factories outperform control factories on all sub-indexes, the largest treatment effect is on the SC responsibilities sub-index.[52] Treatment factories outperform control factories on this sub-index by 0.44 sds (FDR p=0.001). The large, positive effect on this sub-index reflects that treatment factories are more likely to conduct legally-required activities. For example, at the second visit, only 15% of control SCs had conducted a risk assessment, while 56% of treatment SCs had conducted at least one. According to reports by SC presidents, worker representatives, and senior managers, treatment SCs also made more regular reports and recommendations to senior manage-

[49] The cheap and costly compliance results hold if formation requirements are dropped.

[50] The increase in costly compliance also depends on the magnitude of the increase in MNCs' monitoring. Monitoring through the SC Program is not perfect; for example, out of 36 treatment SCs that participated before the second data collection visit, 13 had not conducted a risk assessment. In a couple of cases, the research team determined that the factory falsified the record, and in a few others, that managers had conducted risk assessments, but not the SC. In all 13 cases, the Alliance's records show that the SC had conducted a risk assessment before the visit date.

[51] Treatment and control groups are balanced on all sub-indexes at baseline (OA Table CX).

[52] The lack of an effect on SC formation is perhaps unsurprising in light of the Alliance's engagement with factories prior to their becoming eligible for the SC Program.

ment and followed up on these reports more often.

While the intervention does not significantly affect factories' compliance with the requirement that SCs meet at least once per quarter, it does increase their meeting frequency by 58%, from an average of 1.27 to 1.95 meetings per three months.[53] Further, an exploratory text analysis of SCs' meeting minutes reveals that, despite meeting more often, treatment SCs' meeting minutes are 23% longer (p=0.103) (OA Table CXI). Given that control SCs' meeting minutes are, on average, less than half of a page of text, longer meeting minutes suggest more substantive discussions.

To provide more qualitative insights into changes in SCs' discussions, I examine the meeting minutes' textual content. I prepare the text using standard methods (see Gentzkow, Kelly and Taddy (2019)). I represent discussion content in the form of two-word phrases, known as bi-grams. OA Figure CI presents the top 15 bi-grams for treatment and control groups, respectively, prior to and during the enforcement intervention. Treatment SCs' more substantive discussions are reflected in the count distribution's outward shift. Further, general safety terms, such as "fire safety," disappear from treatment SCs' top bi-grams; they are replaced with more specific terms, such as "risk assessment." Finally, decision terms such as "discussion decision" and "unanimous consent" appear for treatment group minutes but are absent for others.

Together, the compliance results and the analysis of SCs' meeting minutes suggest that the enforcement intervention increases SCs' effort and engagement. This interpretation is further supported by an analysis of the treatment effects on workers' perception of SCs' compliance and effectiveness, which I measure using a pre-specified index. In column (2) of OA Table CIII, the first column shows that the intervention improves workers' perception of SC compliance and effectiveness by about 0.20 sds (p-val=0.112).[54]

5.1.3 The effects of MNC enforcement on safety

The next critical question is whether increased compliance translates into improvements in factory safety. Table III presents the results for the index of safety indicators. Treatment factories' safety indicators improve, on average, between 0.14-0.16 sds relative to controls' (with PDS lasso selected controls, FDR p=0.045). This result provides causal evidence that MNCs' interventions to increase compliance with safety labor laws can improve safety.

Panel B presents the treatment effects for each sub-index. OA Table CX presents baseline balance tests for them. There is one baseline imbalance, which is on worker aware-

[53] Most SCs met with the minimum required frequency, 88% of control SCs and 93% of treatment SCs at the second visit.

[54] OA Table CII presents baseline balance tests for worker secondary outcome variables.

ness of SCs. Workers' awareness at treatment factories is lower, although this difference lessens and is not significant at the 5% level when the outlier factory is dropped. Estimated treatment effects on this sub-index should be interpreted with appropriate caution.

Table III: Treatment effects: Primary outcome index variables

	(1)	(2)	(3)	(4)
Panel A	SC Compliance Index		SC Compliance Index	
	Overall		Cheap	Costly
Treatment	0.203***	0.201***	0.067	0.378***
	(0.0592)	(0.0574)	(0.0574)	(0.0968)
	{0.007}	{0.001}		
	[0.001]	[0.001]	[0.226]	[0.000]
Control Mean	0.046	0.046	0.071	0.014
Observations	80	80	80	80
Strata FE	Y	Y	Y	Y
Control, base. dep. var.	Y	N	Y	Y
PDS Lasso Selected Controls	N	Y	N	N
Panel B	Safety Indicators Index		Job satisfaction and mental well-being Index	
Treatment	0.143**	0.157**	-0.149*	-0.136*
	(0.0674)	(0.0657)	(0.0786)	(0.0714)
	{0.105}	{0.045}	{0.118}	{0.083}
	[0.046]	[0.022]	[0.061]	[0.073]
Control Mean	0.103	0.103	-0.013	-0.013
Observations	80	80	80	80
Strata FE	Y	Y	Y	Y
Control, base. dep. var.	Y	N	Y	N
PDS Lasso Selected Controls	N	Y	N	Y

Notes: This table reports OLS estimates of treatment effects on the primary outcome index variables. Outcome variables are listed at the top of each column. In all cases, higher values of the index correspond to "positive" outcomes. Each column reports the estimated ITT effect from a separate regression. Robust standard errors are reported in round brackets. *p*-values adjusted to control the FDR across primary outcomes are reported in curly brackets. RI *p*-values based on 5000 draws are reported in square brackets. In Panel A, column (2), the PDS lasso selected control variables include the baseline value of the dependent variable and an indicator for having a trade union. In Panel B, column (2), the PDS lasso selected control variables include the square of average years of workers' education and an indicator for having a trade union. In Panel B, column (4), the PDS lasso selected control variables include the square of average years of workers' education and an indicator for having a trade union. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table IV: Treatment effects: Sub-indexes of primary outcome index variables

	(1)	(2)	(3)	(4)	(5)
Panel A: SC Compliance sub-indexes					
	Formation	Operations	Responsibilities		
Treatment	0.0797	0.0549	0.442***		
	(0.0770)	(0.0843)	(0.119)		
	{0.437}	{0.526}	{0.001}		
	[0.298]	[0.522]	[0.000]		
Control Mean	0.118	0.184	-0.149		
Observations	80	80	80		
Strata FE	Y	Y	Y		
Control, base. dep. var.	Y	Y	Y		
Panel B: Safety indicators sub-indexes					
	Spotcheck sub-index	CAP completion sub-variable	Worker SC awareness sub-index	Worker safety knowledge sub-index	Senior manager awareness sub-index
Treatment	0.217**	0.0233	0.197	-0.0943	0.0749
	(0.0892)	(0.0840)	(0.1590)	(0.1377)	(0.2371)
	{0.093}	{1.000}	{0.786}	{1.000}	{1.000}
	[0.015]	[0.794]	[0.189]	[0.503]	[0.805]
Control Mean	0	0.345	0.0490	0.551	0.0860
Observations	80	80	80	80	80
Strata FE	Y	Y	Y	Y	Y
Control, base. dep. var.	Y	Y	Y	Y	Y
Panel C: Workers' job satisfaction and mental well-being sub-indexes					
	Job satisfaction sub-index	Mental well-being sub-index	Turnover sub-variable	Absenteeism sub-variable	
Treatment	-0.386**	-0.0587	-0.0105	-0.0844	
	(0.1565)	(0.1591)	(0.0635)	(0.0547)	
	{0.069}	{0.769}	{0.769}	{0.236}	
	[0.017]	[0.709]	[0.884]	[0.162]	
Control Mean	-0.130	0.0110	0.115	0.0880	
Observations	80	80	80	80	
Strata FE	Y	Y	Y	Y	
Control, base. dep. var.	Y	Y	Y	Y	

Notes: This table reports OLS estimates of treatment effects on sub-indexes of primary outcome index variables. Outcome variables are listed at the top of each column. In all cases, higher values of the index correspond to "positive" outcomes. Each column reports the estimated ITT effect from a separate regression. Robust standard errors are reported in round brackets. *p*-values adjusted to control the FDR across each primary outcome's sub-indexes are reported in curly brackets. RI *p*-values based on 5000 draws are reported in square brackets. *p<0.1; **p<0.05; ***p<0.01.

The first row of Panel B shows that the treatment improves factories' performance on the safety spotcheck conducted by the research team. Treatment factories outperform controls on the safety spotcheck by 0.22 sds (FDR p=0.093). OA Table CXIII shows the treat-

ment effects on each component of the spotcheck index.[55] Treatment factories outperform controls on nearly every component. For example, workers are 9-18% more likely to be found using machines with appropriate guards and to be wearing required personal protective equipment (PPE) for their tasks.[56] Although none of the individual differences is statistically significant, when aggregated, they indicate that the intervention has a small, positive effect on physical indicators of factory safety. This improvement is consistent with the dramatic increase in treatment SCs' implementation of risk assessment. The second column of Table IV Panel B shows that the SC Program does not increase factories' progress on completing their CAPs for building safety violations. There are two likely reasons why. First, 25% of sample factories had completed 90% or more of required remediation actions by baseline. Second, the outstanding violations often required significant financial investment, and even if the MNCs' intervention increases SCs' ability to push for this investments, it may require more time for the effect to materialize.

Finally, the enforcement intervention does not have statistically significant effects on the safety culture sub-indexes (Panel B, columns 3-5). Workers' awareness of SCs increases compared to controls', but the difference is not statistically significant. It is relevant that all sample factories participated in the Alliance's Fire Safety and Worker Helpline Training Program, which included information about the factory's SC and likely helps to explain workers' high baseline level of awareness of SCs and the null result: At baseline, 81% of workers reported being aware of SCs' role and responsibilities, and 89% knew that their factory had a SC. As OA Table CXII shows, though, even with this very high awareness, the intervention increases some measures of workers' awareness.

Visitors to medical clinics: I analyze factories' medical clinic records to test for effects on injuries and illnesses. This outcome is not one of my primary outcomes, and these results should be interpreted as suggestive. Records are available for 62 factories. The records were photographed during onsite visits for later digitization. This process was time intensive, so social compliance assessors were instructed to photograph a sample of three to six days per month when records were large.[57] Due to a misunderstanding about how the records were to be used, assessors sometimes did not always photograph complete daily records. For this reason, and due to the aforementioned legibility issue, the number of days observed per factory varies. To address this issue, I present results with and without probability weights based on the pre-baseline number of days observed.

I focus on the best-measured outcome available, which is the daily count of visitors

[55] Four spotcheck checklist variables drop from the analysis because all factories were found to comply.
[56] PPE includes equipment such as eye guards, finger guards, gloves, goggles, and boots.
[57] Assessors were to photograph the fifth, fifteenth, and twenty-fifth days of month. If these days were weekends or holidays, they were instructed to photograph the preceding and following days.

to the medical clinic. Using monthly data on the number of employees, I calculate the proportion of the workforce that visits the clinic per day. Table II shows that treatment and control factories are balanced on this outcome. To smooth noise in the daily records, to address differences in the number of day-level observations per factory, and to present results at the same level as the business competitiveness outcomes, I average by month.[58]

Panel A of OA Table CXI presents the results. Column (2), which presents weighted results, shows that on average, 1.1% of the workforce visited the medical clinic per day in control factories. In treatment factories, however, the proportion of the workforce visiting the clinic is 15.4% lower (p=0.087). The unweighted estimate is very similar, a 15.5% decrease (p=0.111) (column (3)). The marginal statistical significance may be due to the smaller sample size. Further, in light of the possibility that the intervention may increase workers' willingness to report injuries and illnesses, the estimates may be upward biased. Together, though, the results provide suggestive evidence that enforcing SCs reduces the workforce's need for medical attention by between 15-16%.

5.2 MNC enforcement, workers' well-being, and factories' competitiveness

5.2.1 The effects of MNC enforcement on workers

Panel B of Table III presents the results on workers' job satisfaction and well-being. The estimated treatment effect is -0.14 sds (cols (3)-(4), FDR p=0.083). The estimated effect is unchanged when the outlier factory is dropped (OA Table CVI). Panel C of Table IV displays the estimated effects for the sub-indexes/variables.[59] The results reveal that the negative effect on the primary index is driven by the job satisfaction sub-index (-0.39 sd effect, FDR p=0.075). The estimated treatment effects on mental well-being, turnover, and absenteeism are all negative, but they are smaller in magnitude and are not statistically significant.[60] Consistent with the null effect for turnover, the intervention does not affect workforce composition (OA Table CXII). Together, these results rule out the possibility that changes in workforce composition drive the negative effect.

To further unpack the negative effect on job satisfaction, OA Table CXIV shows the estimated treatment effect on each variable in this sub-index. Panel A shows that the

[58] There is no statistical difference in the average number of observations for treatment or control factories, either at the day- or month-levels (results not reported).

[59] Treatment and control factories are balanced on sub-indexes at baseline (OA Table CX).

[60] For inclusion in the index, the absenteeism and turnover sub-variables are constructed by collapsing five pre- and post-intervention monthly observations into one pre- and post-observation, respectively. They are then multiplied by -1 in order to be unidirectional with other outcomes.

negative effect on job satisfaction is driven by an increase in the proportion of workers considering leaving their factory for safety-related reasons. While the effect is large, a 79% increase, it is on a control group mean of only 2.4% of workers.

Why does MNC enforcement negatively affect some measures of workers' job satisfaction? In Section 5.3, I show that the effect is driven by factories with poor managerial practices where the intervention does not improve compliance or safety. I provide suggestive evidence that it may be due to the intervention raising workers' expectations about what SCs will deliver, and SCs' actual performance not meeting these expectations. I have checked for evidence of other mechanisms, such as workers learning about unsafe conditions. The data do not provide support for a learning mechanism (results not reported).

5.2.2 The effects of MNC enforcement on factories' competitiveness

In this section, I analyze the intervention's effects on labor productivity, gross wages, and employment. There is one control factory that suspended production for three months between the first and second data collection visits. Because this type of temporary suspension is part of business, it does not mean that this factory should be removed from the analysis. But due to the timing of the partial shut down and my smaller sample size, my results may be sensitive to its inclusion.[61] For labor productivity, I present results for the full sample trimming the 1st and 99th percentiles of observations and for the sample dropping the factory that partially shuts down.

Panel A of Table V shows the estimated treatment effects on labor productivity using equation 2.[62] The estimated treatment effects on labor productivity from the trimmed sample range between 8.7-10.6% increases (columns (1)-(3)), although these estimates are not statistically significant. Once the control factory that temporarily shut down is removed, the estimated treatment effects are much smaller, between 3.6-4.6% increases (columns (4)-(6)). None of these estimated effects is statistically significant.

While imprecise, the 95% confidence intervals for these estimates are mostly positive: With the full sample, with 95% confidence, I can rule out effects more negative than -3.3 to -4% (similar when dropping outlier). Together, the results support the interpretation that the intervention did not negatively affect labor productivity; it is much more likely

[61] The factory does not dramatically cut employment, and gross wages do not dramatically fall during the three month period. Employment and wage results are very similar when this factory is dropped. Other main results are also unchanged.

[62] OA Table CIII presents effects on physical output, working hours, and output quality (defects per 100 units); the estimated treatment effects on all three are small and are not statistically significant.

that the effects were zero or positive.[63] In light of the null results, OA Table CXV reports the ex post MDE that would be detectable under standard assumptions for power calculations (80% power and 5% statistical significance level).

Panel B of Table V presents the estimated treatment effect on gross wages and on employment. Columns (1)-(2) show the estimated treatment on wages, which is a -0.30 to -1.5% decrease in wages (not statistically significant). Turning to employment, the estimated treatment is a -0.80 to -1.1% decline in employment (not statistically significant). Again, OA Table CXV reports the MDEs for both variables.

Taken together, the estimated treatment effects suggest that the intervention improves safety without adversely affecting labor productivity, employment, or wages. These results are consistent with the intervention imposing only small costs on suppliers with possible benefits to labor productivity. The results support the potential for MNCs' enforcement interventions to improve labor standards without coming at significant costs in terms of suppliers' efficiency.

5.3 Heterogeneous treatment effects by managerial capacity

In this section, I explore heterogeneity in the intervention's effects depending on factories' baseline managerial capacity.[64] In Section 3, I suggest that more capable managers may have lower marginal costs of compliance, both for cheap and costly requirements or perhaps only for costly ones. If so, then MNC enforcement should cause higher ability managers to increase costly compliance more. Neither high nor low ability managers should increase cheap compliance in response to MNC enforcement.

My main measure of managerial practices summarizes senior and lower-level managers' reported frequency of holding production-related meetings with workers.[65] This question is a variant of questions asked in the World Management Survey (WMS) and in studies on managerial practices by Bloom et al. (2013) and Macchiavello et al. (2015). It measures one specific managerial practice, as it was not feasible to conduct a com-

[63] Treatment effect estimates for all business competitiveness variables are very similar using a panel regression model. Results available upon request.

[64] In my PAP, I specified three other dimensions of heterogeneity to explore: Factory size, compliance with the SC regulation, and location in an EPZ. I find the most compelling pattern of heterogeneous treatment effects for managerial practices, so I present the results for the other dimensions of heterogeneity in the OA Tables DII and DIII. For space reasons, I omit HTE analysis for these dimensions for business competitiveness variables. For location in an EPZ, as there are large differences between the seven treatment and the seven control factories in EPZs (OA Table DII). As such, I depart from the PAP and do not analyze this dimension. Results available upon request.

[65] The measure places 25% weight on the factory's most senior manager's report and 75% weight on the lower-level managers' reports. On average, 15 lower-level managers were surveyed.

Table V: Treatment effects: Business competitiveness outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A	Log(Labor productivity)					
Treatment	0.0915	0.0872	0.1060	0.0387	0.0363	0.0457
	(0.0660)	(0.0605)	(0.0742)	(0.0400)	(0.0364)	(0.0412)
				{0.340}		{0.251}
	[0.145]	[0.179]	[0.104]	[0.374]	[0.408]	[0.317]
Factories	75	75	75	74	74	74
Observations	368	368	368	370	370	370
Stratification variables	Y	Y	Y	Y	Y	Y
Control, baseline dep. var.	Y	Y	N	Y	Y	N
Product FE	N	Y	N	N	Y	N
PDS Lasso Selected Controls	N	N	Y	N	N	Y
Dropping outlier	N	N	N	Y	Y	Y
Panel B	Log(Gross wages)		Log(Employment)			
Treatment	-0.0154	-0.0035	-0.0106	-0.0082		
	(0.0289)	(0.0278)	(0.0213)	(0.0218)		
	{0.450}	{0.669}	{0.450}	{0.669}		
	[0.601]	[0.914]	[0.651]	[0.745]		
Factories	72	72	80	80		
Observations	360	360	400	400		
Stratification variables	Y	Y	Y	Y		
Control, baseline dep. var.	Y	N	Y	N		
PDS Lasso Selected Controls	N	Y	N	Y		

Notes: This table reports OLS estimates of treatment effects on labor productivity, employment, and gross wages. Outcome variables are listed at the top of each column. Each column reports the estimated ITT effect from a separate regression. Panel A reports results for labor productivity. In columns (1)-(3), the sample is trimmed at the 1st and 99th percentile of all factory-month labor productivity observations. In columns (4)-(6), a factory in the control group that partially shut down during the study is dropped. In Panel B, each regression includes five post-treatment observations per factory, where each observation is one month. The regression sample changes across columns due to differential data availability. Standard errors clustered at the factory level are reported in round brackets. *p*-values adjusted to control the FDR across primary outcomes are reported in curly brackets. RI *p*-values based on 5000 draws are reported in square brackets. In Panel A, column (3), the PDS lasso selected control variables include baseline log(labor productivity), its square, an indicator for having a trade union, and indicators for packaging and weaving production. The same controls are selected in column (6), except for baseline log(labor productivity). In Panel B, column (2), the PDS lasso-selected control variables include baseline log(gross wages), its square, the mean log baseline number of employees, and indicators for dyeing, packaging, and weaving production. In column (4), they include baseline log(employment), its square, and indicators for packaging and weaving production. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

plete management diagnostic. As such, one may question whether this measure reflects broader managerial capacity. OA Figure CII provides evidence that it does. It presents a binned scatterplot that includes all apparel manufacturers from all countries included in the WMS. It shows that apparel firms' score on the WMS's meeting question is highly correlated with their average overall WMS Index excluding this question.[66] Evidently, this question captures meaningful information about firms' overall managerial practices.

I partition the sample into above/below median groups using the baseline value of managerial practices. I refer to the below median group as poorly-managed and to the above median group as better-managed. OA Table CXVI shows baseline balance within each interaction-term subgroup for primary outcome variables for non-attrited factories. There are no statistically significant differences between subgroups.

Panel A of Table VI presents the main results. Each column considers a different primary outcome variable. The first row displays the estimated treatment effect for the below median group, and the second row displays the treatment effect for the above median group. The final row displays the p-value of the difference in the treatment effects on the subgroups. The regression specification is equation 3.

Beginning with SC compliance (column (1)), the estimated treatment effect for poorly-managed factories is 0.11 sds (not statistically significant). In contrast, the estimated effect for better-managed factories is 0.31 sds ($p \approx 0.000$). The difference between these estimates is statistically significant (p=0.069). OA Table **??** further examines heterogeneous effects on cheap and costly compliance. Consistent with the model, neither poorly-managed nor better-managed factories increase cheap compliance (column (1)). Better-managed factories, however, dramatically increase costly compliance. In contrast, for poorly-managed factories, while the point estimates are qualitatively larger than those for cheap compliance, costly compliance does not significantly improve.

Returning to Table VI, for poorly-manged factories, the small improvement in compliance translates into a very small improvement in safety indicators (0.05 sd effect, p=0.540). In contrast, consistent with the large effect on compliance, better-managed factories improve safety indicators by 0.24 sds (p=0.027). The difference between the estimates is not statistically significant. Together, though, the results suggest that the enforcement intervention leads to large improvements in factories' compliance and in SCs' effectiveness, but only among factories with better baseline managerial practices.

Turning to column (3), again, there is a stark difference in the estimated effects on job satisfaction and mental well-being for poorly- and better-managed factories. The esti-

[66] The WMS question asks whether performance is reviewed with appropriate frequency and communicated to staff ((World Management Survey, n.d.)). The index is the average score on all other questions.

Table VI: Heterogeneous treatment effects by managerial practices

	SC Compliance	Safety Indicators	Job Satisfaction & Mental Well-being
	(1)	(2)	(3)
Panel A: Treatment Phase			
Below median	0.105	0.050	-0.271**
	(0.0823)	(0.0819)	(0.1252)
	[0.230]	[0.566]	[0.027]
Above median	0.311***	0.235**	-0.030
	(0.000)	(0.1039)	(0.1067)
	[0.000]	[0.049]	[0.754]
p-val, diff	0.069	0.165	0.162
	[0.133]	[0.239]	[0.145]
Observations	80	80	80
Panel B: Post-treatment phase			
Below median	0.127	0.031	0.009
	(0.1247)	(0.1162)	(0.1340)
	[0.333]	[0.789]	[0.941]
Above median	0.292***	0.090	0.240*
	(0.1058)	(0.0857)	(0.1269)
	[0.016]	[0.307]	[0.065]
p-val, diff	0.306	0.692	0.225
	[0.364]	[0.702]	[0.216]
Observations	80	80	80
Panel C: Pooled			
Below median	0.116	0.041	-0.131
	(0.0757)	(0.0699)	(0.0918)
	[0.195]	[0.613]	[0.216]
Above median	0.302***	0.162**	0.105
	(0.0660)	(0.0651)	(0.0912)
	[0.001]	[0.084]	[0.239]
p-val, diff	0.060	0.208	0.075
	[0.173]	[0.332]	[0.093]
Observations	160	160	160
Factories	80	80	80
Stratification variables	Y	Y	Y
Control, base. dep. var.	Y	Y	Y

Note: This table reports OLS estimates of heterogeneous treatment effects on primary outcome index variables. Each outcome variable is indicated at the top of the table. Each panel reports the results for data collection during the treatment phase, after the treatment phase, and pooling both rounds of data collection. In each panel, the "Below median" row reports the estimated treatment effect for the subgroup with below median baseline managerial practices. In each panel, the "Above median" row reports the estimated treatment effect for the subgroup with above median baseline managerial practices. The final two rows in each panel report the *p*-value of the difference between the estimated treatment effects for below and above median subgroups. RI *p*-values based on 5000 draws are reported in square brackets. All regressions include stratification variables and a control for the baseline value of the dependent variable. All subgroups have 40 observations. Robust standard errors are reported in round brackets. *p<0.1; **p<0.05; ***p<0.01.

mated treatment effect for poorly-managed factories is -0.27 sds (p=0.034). The estimated effect for better-managed factories, though, is close to zero (-0.030 sd effect, p=0.778). The p-value for the difference in the treatment effects is p=0.162. Evidently, the decline in job satisfaction found in Section 5.2.1 is driven by poorly-managed factories, for which the intervention does not meaningfully improve compliance or safety.

I can try to increase statistical power to detect differences between the treatment effects for poorly- and better-managed factories by pooling the treatment and post-treatment rounds of data. Panel C displays the results. Column (1) shows that the estimated treatment effects on compliance for above and below median managerial practice factories remain stable. The difference between these estimates is statistically significant (p=0.060). Turning to column (2), the estimated treatment effects on safety indicators, although attenuated, exhibit the same pattern. I am unable, however, to reject that the effects for both groups are equal (p=0.208). Finally, in column (3), the difference in effects is fairly stable, and I reject the null of equality with p=0.075. In sum, while not conclusive for the safety indicators index, the results provide support for the MNCs' enforcement having differential effects on poorly- and better-managed factories.

Next, I examine whether the pattern of heterogeneous treatment effects is present for visitors to factories' medical clinics. Panel B of OA Table CXI presents the results. In both specifications, the pattern of estimates is consistent with the decline in visitors being driving by better-managed factories. While I am underpowered to detect differences in the treatment effects, these results are consistent with the other evidence presented on the important role of managerial practices.

Finally, I test for heterogeneous effects on business outcomes. OA Table CXVII presents the results. There is not strong evidence of heterogeneous effects, which allays concerns that the null main effects mask negative treatment effects for better-managed factories.

In OA D, I present robustness checks for the HTE analysis that also support these results. Specifically, I show that correlation between factories' management practices and other characteristics does not drive the results. I allay concerns that differential monitoring may generate the heterogeneity. Finally, I implement the analysis using an alternative measure of management practices and find qualitatively similar results.

To summarize, I find evidence that managerial capacity plays an important role in determining the MNCs' enforcement intervention effects. The intervention improves compliance and safety indicators only in factories with better managerial capacity. These improvements do not come at the cost of negative effects on workers' job satisfaction and well-being. For factories with poor management practices, however, the intervention does not improve compliance or safety-related outcomes and has a negative effect on in-

dicators of workers' job satisfaction.

Why do indicators of job satisfaction decline at poorly-managed treatment factories?

Why are indicators job satisfaction at poorly-managed factories negatively affected when the intervention has little to no effect on compliance and safety indicators? One plausible mechanism is that the MNCs' intervention raises workers' expectations for SCs, but in poorly-run factories, workers' expectations are not met, and they are disappointed. As discussed in Section 1, this effect would be consistent with recent experimental findings by Adhvaryu, Nyshadham and Xu (2018).

I cannot directly test for this possibility, as I did not collect data on workers' expectations for SCs. I find support for an important role for workers' expectations and for learning, though, from qualitative evidence from eight interviews with compliance managers at treatment factories.[67] Multiple managers reported that it took several months after their factory's SC became active for workers to understand what issues they could report to the SC and expect to have resolved. In particular, managers reported that it was initially common for workers to raise issues to the SC that were outside of its authority (e.g., working hours or wage-related concerns). If these concerns went unresolved, workers may have been disappointed. It is also unsurprising that workers may not initially understand the scope of SCs' authority, as the SC Program is likely the first time that they have been told that there is a democratically-selected institution at the factory that is responsible to address their concerns.[68] I do find that workers' perception of SCs' effectiveness is unchanged in poorly-managed factories and only improves in better-managed factories (results not reported). If unmet expectations play a role, it suggests that the negative effect may be temporary. I test this possibly in the next section.

5.4 Do the effects persist after MNCs cease intensive enforcement?

The Alliance's SC Program aims to bring factories into compliance through intensive enforcement for a period of six months. The Alliance then continues to monitor factories under its general monitoring activities. Do factories maintain improvements in compliance and safety indicators after the MNCs cease intensive enforcement? Table VII presents the

[67] I selected two treatment factories per quartile of management practice scores for interviews. Selection was not done randomly but based on factories' proximity to Dhaka and managers' availability.

[68] I do not have data with which I can test whether workers raise non-safety issues to SCs. I have access to records of issues raised to SCs, but SCs only recorded the safety-related issues in these documents. Further, I did not ask workers about the specific issues that they raised to the SC.

estimated treatment effects on the first three primary outcomes measured 3-4 months after the end of the intensive enforcement period for treatment factories. Column (1) shows that the estimated treatment effect on compliance persists. The effect remains around 0.21 sds (FDR p=0.078). Columns (2)-(3) show that the persistence is driven by continued outperformance on the costly compliance index; in contrast, the effects on cheap compliance remain small and not statistically significant.

While treatment factories continue to significantly outperform controls on the compliance index, they no longer do so on the safety indicators index. Treatment factories perform 0.06-0.07 sds better than controls. The dissipation is partially due to a slight attenuation in treatment factories' performance, but control factories also improve their performance. I argue that the convergence is likely due to control factories' expectations about future enforcement by the Alliance. The Alliance rolls out its programs in a staggered fashion, and factory managers generally know that they will be required to participate. For example, in a survey that I conducted with compliance managers of factories under shared-ownership with experimental factories, 46% of managers report taking actions to prepare for the SC Program.[69] Further, Figure III shows that while control factories' engagement in cheap compliance remains fairly flat between the second and third data collection rounds, they begin to increase adoption of costly compliance.

Interestingly, the negative effect on the job satisfaction and mental well-being index disappears, and the estimated treatment effect is actually positive (not statistically significant). Panel B of Table VI shows that the effect for poorly-managed factories is zero, while the effect for better-managed factories is 0.24 sds (p=0.065), although the estimates are not statistically different. This pattern of results is consistent with the possibility that the short-run negative effect for poorly-managed factories is due to disappointment. For better-managed factories, the positive estimate is consistent with workers' valuing SCS, but it requiring time for them to learn about SCs' role or for SCs' benefits to materialize.

Together, the results suggest that under less intensive monitoring, treatment factories maintain compliance improvements. Less observable improvements in safety indicators, however, attenuate slightly. Control factories, possibly expecting future enforcement by the MNCs, begin to improve compliance and safety indicators. Treatment factories' maintenance of costly compliance improvements is consistent with their subjection to continued audits by the Alliance: 15% were audited in the six months after completing it.

Finally, I do not find delayed adverse effects on labor productivity, employment, or

[69]The Alliance also agreed to delay controls' participation in the SC Program until the end of the data collection period. No control factories were treated during the experiment, but it's possible that Alliance personnel communicated to control factories that they would eventually be required to participate.

wages. Estimates from my preferred specifications are close to zero. See OA Table CXVIII.

Table VII: Persistence of treatment effects after end of intensive enforcement: Primary outcome index variables

	SC Compliance Overall	SC Compliance Cheap	SC Compliance Costly	Safety Indicators	Job Satisfaction & Mental Well-being
	(1)	(2)	(3)	(4)	(5)
Treatment	0.213**	0.087	0.375***	0.063	0.115
	(0.0832)	(0.0786)	(0.1216)	(0.0692)	(0.0909)
	{0.078}			{1.000}	{1.000}
	[0.012]	[0.279]	[0.003]	[0.371]	[0.206]
Control Mean	0.149	0.094	0.219	0.156	-0.099
Observations	80	80	80	80	80
Strata FE	Y	Y	Y	Y	Y
Control, base. dep. var.	Y	Y	Y	Y	Y

Notes: This table reports OLS estimates of the persistence of treatment effects on primary outcome index variables measured 3-4 months after the end of the intensive enforcement period. Outcome variables are listed at the top of each column. In all cases, higher values of the index correspond to "positive" outcomes. Each column reports the estimated ITT effect from a separate regression. Robust standard errors are reported in round brackets. *p*-values adjusted to control the FDR across primary outcomes are reported in curly brackets. RI *p*-values based on 5000 draws are reported in square brackets. *p<0.1; **p<0.05; ***p<0.01.

5.5 External Validity

In this Section, I discuss the extent to which this study's results generalize to other settings. I identify four main threats to external validity: First, one may be concerned that the SC Program occurred in a context in which many other safety initiatives were ongoing and that this may not be the case in other contexts in which MNCs enforce local labor laws. Second, one may be concerned that coalition-based enforcement approaches are not common. Third, one may be concerned that SCs are a unique feature of Bangladeshi labor law. Finally, one may be concerned that Bangladesh's apparel sector is somehow unique.

First, as detailed in Section 2.1, MNCs monitor social and environmental performance throughout their supply chains. Often, Western MNCs' supplier bases overlap, and suppliers are monitored through multiple programs.[70] For example, from 2017-2018, condi-

[70]In fact, a literature in operations management examines the implications of the proliferation of supplier codes of conduct and social and environmental compliance monitoring regimes; it advocates for more standardized approaches (Busse, 2010, Schleper and Busse, 2013).

tional on appearing in the Alliance's supplier base, a factory supplied an average of 1.7 members. Alliance members sourced from and very likely monitored suppliers in other countries; for example, from 2015-2019, Gap Inc. conducted fewer than 10% of its supplier audits in Bangladesh (Amengual and Distelhorst, 2019).[71] In short, in GVCs with Western MNC buyers, participation in social compliance enforcement programs is part of doing business for upstream firms.

Second, coalition-based approaches are a common way of organizing enforcement in GVCs. For example, the Life and Building Safety Initiative, which includes several former Alliance members, enforces building safety standards in the apparel and homegood sectors in India and Vietnam. The Action for Living Wages Initiative enforces FOA rights and wages in 21 apparel, textile, and footwear MNCs' (e..g., H&M, Inditex, Primark) supplier bases in Bangladesh, Cambodia, Myanmar, and Turkey. In Myanmar, its members signed guidelines for FOA rights with labor unions and employers; the MNCs' commitment entails terminating business relations with suppliers that do not comply with the guidelines. Evidently, the coalition-based enforcement, the threat of suspension, and the broader trade-offs that the suppliers in my study face are similar to those in other settings.

Third, as I discuss in Section 2.4, SCs are a core component of international OSH standards and feature in many countries' OSH laws. Finally, the apparel sector is common to many developing countries, as it is both low-skill and labor-intensive. And as discussed in a recent ILO report on working conditions in GVCs, the labor issues that exist in Bangladesh's apparel sector are common in apparel-producing countries (ILO, 2016).

6 Conclusion

In this paper, I analyze the effects of a coalition of MNCs' CSR program to enforce a local labor law on their Bangladeshi suppliers. This study is a "first" in multiple streams of literature. It is the first study to provide experimental evidence on whether firms' CSR programs generate social benefits, which in this case is improved labor standards in GVCs. It is also the first study to provide experimental evidence on the effects of enforcing labor regulation on factories' competitiveness and workers' well-being. Further, it is the first study to experimentally intervene to increase collective worker voice in the firm.

I find that the MNCs' enforcement intervention is successful at increasing factories' compliance with a Bangladeshi labor law that mandates SCs and that it improves indicators of factory safety. The results demonstrate that MNC enforcement can cause factories

[71] Because the Alliance's remit was focused on building and fire safety, the Alliance's monitoring activities would not substitute for broader social compliance programs, which cover many other areas.

to go beyond box ticking and to increase meaningful compliance with the law. I also find, however, that suppliers do not achieve full compliance. My findings are consistent with a model in which MNCs provide positive but relatively smaller penalties for noncompliance. MNCs may opt *not* to set penalties high enough so as to induce full compliance.

I do not find evidence of negative effects on labor productivity, employment, and wages. Together, these results help to allay concerns that enforcement of labor regulation necessarily entails trade-offs between competitiveness and improved working conditions. Further, they can help economists to update their views on enforcement of labor regulation and economic outcomes in developing countries.

Finally, I find that providing firms with incentives to comply with the SC law is not sufficient to increase compliance but that firms' capacity to respond to incentives also matters. In particular, the estimated treatment effects on compliance and safety are large for better-managed factories but are small for poorly-managed factories.

My findings raise several directions for future research. First, this research suggests that *private* sector capacity plays an important role in determining the efficacy of *public* regulation. Future research can investigate how firms' capacity supports compliance. Second, in the short-run, I do not find evidence of workers differentially sorting in response to improvements in firms' compliance. It is possible, though, that if improvements are sustained, it may affect workers' mobility and sorting into factories. Boudreau, Heath and McCormick (2020) provide evidence that garment workers who begin their careers with poor information about factories' working condition exhibit a revealed preference for improving their working conditions compared to their wages. More research is needed, however, on how workers in developing countries trade-off between wages and non-pecuniary amenities. Third, a critical question is what the general equilibrium effects of multinational enforcement of labor law are on compliance and competitiveness of the targeted sector. Finally, there is generally a dearth of empirical evidence in economics on the welfare effects of firms' CSR activities, which are becoming increasingly common and large-scale. These interventions merit more attention.

References

Adhvaryu, Achyuta, Anant Nyshadham, and Huayu Xu. 2018. "Hostel Takeover: Living Conditions, Reference Dependence, and the Well-being of Migrant Workers." *Working Paper*.

Aghion, Philippe, Robin Burgess, Stephen J Redding, and Fabrizio Zilibotti. 2008. "The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India." *The American Economic Review*, 98(4): 1397–1412.

Alliance. 2013. "Members Agreement."

Alliance. n.d.. "Safety Committees." *http://bangladeshworkersafety.org/what-we-do/worker-empowerment/sc*, Accessed: 2020-12-04.

Almunia, Miguel, Jonas Hjort, Justine Knebelmann, and Lin Tian. 2020. "Strategic or Confused Firms? Evidence from "Missing" Transactions in Uganda." *Working Paper*.

Amengual, Matt, and Greg Distelhorst. 2019. "Can Sourcing Help Enforce Global Labor Standards? Evidence from the Gap Inc Supply Chain." *Working Paper*.

Amirapu, Amrit, and Michael Gechter. 2020. "Indian Labor Regulations and the Cost of Corruption: Evidence from the Firm Size Distribution." *Review of Economics and Statistics*.

Anderson, Michael 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.

Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack. 2014. "No margin, no mission? A field experiment on incentives for public service delivery." *Journal of Public Economics*, 120: 1–17.

Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S Lee. forthcoming. "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the delivery of Public Services." *American Economic Review*.

Athey, Susan, and Guido Imbens. 2016. "The Econometrics of Randomized Experiments."

Bai, Jie, Ludovica Gazze, and Yukun Wang. 2019. "Collective Reputation in Trade: Evidence from the Chinese Dairy Industry." *NBER Working Paper No. 26283*.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy*, 169–217.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on Treatment Effects after Selection among High-Dimensional Contr." *Review of Economic Studies*, 81(2): 608–650.

Bénabou, Roland, and Jean Tirole. 2010. "Individual and Corporate Social Responsibility." *Economica*, 77: 1–19.

Besley, Timothy, and Maitreesh Ghatak. 2007. "Retailing public goods: The economics of corporate social responsibility." *Journal of Public Economics*, 91: 1645–1663.

Besley, Timothy, and Robin Burgess. 2004. "Can Labor Regulation Hinder Economic Performance? Evidence from India." *The Quarterly Journal of Economics*, 119(1): 91–134.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *The Quarterly Journal of Economics*, 128(1): 1–51.

Botero, Juan C, Simeon Djankov, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2004. "The Regulation of Labor." *The Quarterly Journal of Economics*, 119(4): 1339–1382.

Boudreau, Laura, Rachel Heath, and Tyler McCormick. 2020. "Migrants, Information, and Working Conditions in Bangladeshi Garment Factories." *Working Paper*.

Burbano, Vanessa C. 2016. "Social Responsibility Messages and Worker Wage Requirements: Field Experimental Evidence from Online Labor Marketplaces." *Organization Science*, 27(4): 1010–1028.

Busse, Christian. 2010. "Codes of Conduct in Organisational Context: From Cascade to Lattice-Work of Codes." *Journal of Business Ethics*, 94: 471–487.

Bustillo, Miguel, Tom Wright, and Shelly Banjo. 2012. "For Wal-Mart, Sears, Tough Questions in Bangladesh Fire." *The Wall Street Journal*.

Callen, Michael, Saad Gulzar, Ali Hasanain, Yasir Khan, and Arman Rezaee. 2018. "Strengthening state capabilities: The role of financial incentives in the call to public service." *NBER Working Paper No. 21180*.

Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." *Quarterly Journal of Economics*, 127: 1755–1812.

Central Intelligence Agency. 2016. "The World Factbook." Central Intelligence Agency.

Compa, Lance. n.d.. "Labor unions can save lives, journal=The Washington Post, year=2013, month=May, day=27."

Dal Bó, Ernesto, Frederico Finan, and Martín A Rossi. 2013. "Strengthening state capabilities: The role of financial incentives in the call to public service." *The Quarterly Journal of Economics*, 128(3): 1169–1218.

de Bakker, Frank, Andreas Rasche, and Stefano Ponte. 2019. "Multi-Stakeholder Initiatives On Sustainability: A Cross-Disciplinary Review And Research Agenda For Business Ethics." *Business Ethics Quarterly*, 29.

Deloitte. 2015. "Global Powers of Retailing 2015: Embracing Innovation." London, United Kingdom.

DiNardo, John, and David S Lee. 2004. "Economic Impacts of New Unionization on Private Sector Employers: 1984–2001." *The Quarterly Journal of Economics*, 119(4): 1383–1441.

Dragusanu, Raluca, and Nathan Nunn. 2018. "The Effects of Fair Trade Certification: Evidence from Coffee Producers in Costa Rica." *Working Paper*.

Dragusanu, Raluca, Daniele Giovannucci, and Nathan Nunn. 2014. "The Economics of Fair Trade." *Journal of Economic Perspectives*, 28: 217–236.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013. "Truth-Telling by Third Party Auditors and the Response of Polluting Firms: Experimental Evidence from India." *The Quarterly Journal of Economics*, 128(4): 1499–1545.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2014. "The Value of Regulatory Discretion: Estimates from Environmental Inspections in India." *Working Paper*.

Fishback, Price V, and Shawn E Kantor. 1996. "Did Workers Pay for Workers Compensation Laws?" *The Quarterly Journal of Economics*, 110(3): 713–742.

Fisman, Raymond, and Yongxiang Wang. 2015. "The Mortality Costs of Political Connections." *Review of Economic Studies*, 82: 1346–1382.

Friedman, Milton. 1970. "The Social Responsibility of Business is to Increase its Profits." *The New York Times Magazine*, 1–19.

Gentzkow, Matthew, Bryan T Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature*, 57(3): 535–574.

Government of Bangladesh. 2015. "Bangladesh Gazette: Ministry of Labor and Employment Notification." *English translation*, S.R.O. No. 291.

Greenhouse, Steven. 2013*a*. "Obama to Suspend Trade Privileges with Bangladesh." *The New York Times*.

Greenhouse, Steven. 2013*b*. "U.S. Retailers See Big Risk in Safety Plan for Factories in Bangladesh." *The New York Times*.

Hainmueller, Jens, Michael J Hiscox, and Sandra Sequeira. 2015. "Consumer demand for fair trade: Evidence from a multistore field experiment." *The Review of Economics and Statistics*, 97: 242–256.

Harrison, Ann, and Jason Scorse. 2010. "Multinationals and anti-sweatshop activism." *The American Economic Review*, 100(1): 247–273.

Hart, Oliver, and Luigi Zingales. 2017. "Companies Should Maximize Shareholder Welfare Not Market Value." *Journal of Law, Finance, and Accounting*, 2: 247–274.

Haushofer, Johannes, and Jeremy Shapiro. 2016. "The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya." *Quarterly Journal of Economics*, 1973–2042.

Heath, Rachel, and Ahmed Mushfiq Mobarak. 2015. "Manufacturing growth and the lives of Bangladeshi women." *Journal of Development Economics*, 115: 1–15.

Heß, Simon. 2017. "Randomization inference with Stata: A guide and software." *Stata Journal*, 17(3): 630–651.

ILO. 1981. "Occupational Safety and Health Convention, 1981 (No. 155)." https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:::NO:12100:P12100_ILO_CODE:C155:NO.

ILO. 2016. "Sectoral Studies on Decent Work in Global Supply Chains." International Labor Organization.

ILO. 2017. "Towards safer working conditions in the Bangladesh ready-made garment sector." International Labor Organization.

Jäger, SImon, Benjamin Schoefer, and Jörg Heining. 2020. "Labor in the Boardroom." *Working Paper*.

Kitzmueller, Markus, and Jay Shimshack. 2012. "Economic Perspectives on Corporate Social Responsibility." *Journal of Economic Literature*, 50: 51–84.

La Porta, Rafael, Florencio Lopez-de Silanes, Andrei Schleifer, and Robert Vishny. 1999. "The Quality of Government." *Journal of Law, Economics, & Organizations*.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies*, 76(3): 1071–1102.

Lee, David S., and Alexandre Mas. 2012. "Long-Run Impacts of Unions on Firms: New Evidence from Financial Markets, 1961–1999 *." *The Quarterly Journal of Economics*, 127(1): 333–378.

Locke, Richard M. 2013. *The Promise and Limits of Private Power: Promoting Labor Standards in a Global Economy.* Cambridge, United Kingdom:Cambridge University Press.

Macchiavello, Rocco, and Ameet Morjaria. 2015. "The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports." *American Economic Review*, 105: 2911–2945.

Macchiavello, Rocco, and Josepa Miquel-Florensa. 2019. "Buyer-Driven Upgrading in GVCs: The Sustainable Quality Program in Colombia." *Working Paper*.

Macchiavello, Rocco, Andreas Menzel, Atonu Rabbani, and Christopher Woodruff. 2015. "Challenges of Change: An Experiment Training Women to Manage in the Bangladeshi Garment Sector." *Working Paper*.

Manning, Alan. 2003. *Monopsony in Motion : Imperfect Competition in Labor Markets.* Princeton University Press.

McKinsey & Company. 2011. "Bangladesh's ready-made garments landscape: The challenge of growth." McKinsey & Company.

Ministry of Labour and Employment. 2013. "National Tripartite Plan of Action on Fire Safety and Building Integrity in the Ready-Made Garment Sector in Bangladesh." Ministry of Labour and Employment, Dhaka, Bangladesh.

Munni, Monira. 2015. "Owners-workers' row delays rules forming." *The Financial Express*.

Munni, Monira. 2017. "Half of apparel units flout needs for safety committees." *The Financial Express*.

O'Brien, Peter C. 1984. "Procedures for Comparing Samples with Multiple Endpoints." *Biometrics*, 40(4): 1079–1087.

O'Rourke, Dara. 2014. "The Science of Sustainable Supply Chain." *Science*, 344(6188): 1124–1127.

Schleper, Martin c, and Christian Busse. 2013. "Toward a standardized supplier code of ethics: Development of a design concept based on diffusion of innovation theory." *Logistics Research*, 6(4): 187–216.

Solidarity Center. 2016. "Bangladesh Garment and Textile Factory Safety Incidents Since Tazreen Fashion Factory Fire, November 24, 2012." Solidarity Center.

Tanaka, Mari. forthcoming. "Exporting Sweatshops? Evidence from Myanmar." *Review of Economics and Statistics*.

World Bank. 2020. "World Development Report 2020: Trading for Development in the Age of Global Value Chains." World Bank, Washington DC.

World Management Survey. n.d.. "World Management Survey Manufacturing Instrument." https://worldmanagementsurvey.org/survey-data/methodology/.

World Trade Organization. 2017. "World Trade Statistical Review 2017." World Trade Organization, Geneva, Switzerland.

Yao, Yang, and Ninghua Zhong. 2013. "Unions and Workers' Welfare in Chinese Firms." *Journal of Labor Economics*, 31(3): 633–667.

Yassi, Annalee, Karen Lockhart, Mona Sykes, Brad Buck, Bjorn Stime, and Jerry M Spiegel. 2013. "Effectiveness of Joint Health and Safety Committees: A Realist Review." *American Journal of Industrial Medicine*, 56: 424–438.

Online Appendices

A: Theoretical Framework

Proof of Proposition 1: The manager chooses the optimal levels of k_1 and k_2 in order to maximize equation 1. If $\theta = 0$, then the manager will set $k_1 = 0$ because his utility is strictly decreasing in k_1. The optimal level of k_2 is: $k_2^* = \frac{p}{c_2}$. If $p < c_2$, then the solution is interior (Case 1.1). Otherwise, $k_2 = 1$ (Cases 2.1 and 3.1). If $\theta \in (0,1)$, then the solution is interior and the optimal levels of k_1 and k_2 are: $k_1^* = \frac{\theta p}{c_1}$ and $k_2^* = \frac{p}{c_2}$ (Case 1.2). If $p \geq c_2$, $k_2^* = 1$ (Cases 2.2, 3.2, and 3.3). k_1^* is interior if $p \leq c_1$ (Cases 1.2 and 2.2). If $p > c_1$, then by comparing the numerator and denominator of k_1^*, we see that it is interior if $\theta < \frac{c_1}{p}$, otherwise, $k_1^* = 1$ (Cases 3.2 and 3.3). □

Corollary 1. (i) The manager's optimal levels of formal and cheap compliance are increasing in p and decreasing in c_1 and c_2, respectively. (ii) The manager's optimal level of costly compliance is increasing in θ, but their optimal level of cheap compliance does does not depend on θ. (iii) The manager's optimal level of cheap compliance is always weakly greater than their optimal level of costly compliance. If penalties are monitoring are not too high ($p < c_1$ or $p > c_1$ and $\theta < \frac{p}{c_1}$), then it is strictly greater.

Heterogeneous managerial ability: Higher ability managers may have a lower marginal cost of compliance. Suppose that managers' cost function now takes the form $\frac{c}{\lambda}$, where $\lambda > 1$ is a manager's type, and higher ability managers have higher values of λ. First, suppose that managerial ability matters for both cheap and costly compliance:

$$\max_{k_1,k_2} V(k_1,k_2) = -\frac{ck_1^2}{2\lambda} - \frac{c_2 k_2}{2\lambda} - p(1-k_2) - \theta\, p(1-k_1) \tag{4}$$

When $\theta = 0$, $k_2^* = \frac{\lambda p}{c_2}$. For given values of p and c_2, higher ability managers engage in more cheap compliance. When $\theta > 0$, the new interior solution is $k_1^* = \frac{\lambda\, \theta\, p}{c_1}$ and $k_2^* = \frac{\lambda p}{c_2}$. For given values of θ, p, c_1, and c_2, higher ability managers engage in more cheap and costly compliance. Finally, note that the comparative static for k_1 and θ, $\frac{\partial k_1^*}{\partial \theta} = \frac{\lambda p}{c_1} > 0$ is positive and is increasing in λ. k_2^* does not depend on θ, so the comparative static is zero.

It's also plausible that managerial skill only matters for costly compliance: It does not require high ability to write down workers' names on a list of SC members, but it may require high ability to correctly organize the democratically selection process. In the extreme, suppose that it only matters for costly compliance. Then the second term of equation 4 becomes $\frac{c_2 k_2^2}{2}$. In this case, when $\theta = 0$, there is no difference in cheap compliance among managers with differing abilities. When $\theta > 0$, the new interior solution is $k_1^* = \frac{\lambda\, \theta\, p}{c_1}$ and $k_2^* = \frac{p}{c_2}$. Higher ability managers engage in more costly compliance, but not more cheap compliance. Finally, note that the comparative static for k_1 and θ,

$\frac{\partial k_1^*}{\partial \theta} = \frac{\lambda p}{c_1} > 0$ is positive and is increasing in λ. k_2^* does not depend on θ, so the comparative static is zero.

B: Truthful Reporting

In any study of compliance, one must be concerned about subjects' incentives to misreport. In the context of a RCT, in order to bias estimates of treatment effects, the treatment would need to affect subjects' propensity to tell the truth. In designing this RCT, I was keenly aware of these concerns, and I strove to design the data collection protocols in order to minimize experimenter demand effects and the potential for the SC Program to affect reporting. In this Appendix, I report results for empirical tests of truth-telling by factories. In the Supplementary Materials, I provide a detailed overview of how the data collection protocol was designed to minimize experimenter demand effects.

During onsite visits, the research team collected data about other Alliance programs. In addition to shielding my interest in SCs, this approach allows me to test for effects on truth-telling and on "placebo" outcomes that I do not expect to be affected be the SC Program. Beginning with senior managers, I asked them questions about their factories' progress with building safety remediation under their Alliance CAP. I also asked about their awareness of the Alliance's worker helpline, including the number of recent reports about their factory to the helpline. I can verify the correct answers to these questions using the Alliance's records. Thus, they allow me to test for non-truthful reporting and for managers' awareness of their factories' safety performance.

I also test for effects on three "placebo" outcomes related to factories' compliance with other Alliance programs. First, the Alliance required that all factory personnel carry its worker helpline phone number card with their employee ID card. Survey enumerators were required to verify that survey participants matched the list of randomly-selected participants, which they did by checking the participant's ID card. While checking, they noted whether the participant carried the helpline card (without indicating this to the survey participant). Thus, I can test whether treatment factories differentially respond to being visited by the research team by increasing the share of personnel wearing the cards. I test for effects for workers and for lower-level managers. Second, I test for effects on factories' maintenance of records of Alliance fire safety training implementation. The Alliance used a "train-the-trainer" model and required factories to conduct periodic training with workers and to maintain a training record using a provided template.

Table BI presents baseline balance for truth-telling variables. In Panel A, variables based on the senior manager survey, there is an imbalance on one variable: Under-reporting of calls to the Alliance worker helpline by senior managers. It is important to note, though, that senior managers at 19 control and 13 treatment factories reported not knowing or were unaware of the Alliance's worker helpline at baseline. In Panel B, there is a marginally statistically significant difference for the share of workers with the Alliance's worker helpline card. This difference shrinks and is no longer statistically significant if the outlier factory on worker variables is dropped.

Table BII reports the results. Beginning with Panel A, columns (1)-(2) report treatment

effects on truth-telling. In column (1), the estimated treatment effect on over-reporting the factory's progress with required building safety remediation is close to zero and not statistically significant. In column (2), managers at treatment factories are actually less likely to under-report calls to the Alliance helpline (not statistically significant). While the treatment does not affect managers' propensity to misreport, columns (3) and (4) show that it appears to increase their awareness of safety issues: Treatment senior managers are significantly more likely to accurately report whether their factory was recently audited by the Alliance on building safety. They are also marginally more likely to be aware of the existence of the Alliance's worker helpline. These findings are consistent with the SC intervention improving senior managers' information - for example, through SCs providing more reports - but not altering their incentives to misreport.

Turning to Panel B, columns (1)-(2) show that there is no difference between treatment or control factories in the share of workers or managers found carrying the Alliance helpline card. Column (3) shows that there is no difference on the Alliance's requirement to maintain safety training records, although compliance with this requirement was already very high at baseline. Together, the results do not provide any evidence that treatment factories differentially respond to the data collection.

Table BI: Baseline balance tests, truth-telling

	Control mean (SD)	T-C diff	*p*-value [RI *p*]	Number of factories
	(1)	(2)	(3)	(4)
Panel A: Senior Managers				
Over-reports CAP completion	0.275	0.006	0.951	77
	(0.4522)		[0.957]	
Under-reports Alliance helpline calls†	0.478	-0.348***	0.006	50
	(0.5107)		[0.007]	
Correctly reports whether CAP visit	0.195	-0.082	0.299	80
	(0.4012)		[0.362]	
Aware of Alliance helpline	0.927	0.047	0.337	80
	(0.2637)		[0.615]	
Panel B: Compliance with other Alliance Programs				
Share workers with helpline card	0.821	-0.103*	0.086	84
	(0.2252)		[0.087]	
Share lower-level managers	0.733	-0.089	0.180	84
with helpline card	(0.3126)		[0.194]	
Alliance Safety Training Record	0.977	0.001	0.977	84
	(0.1525)		[1.000]	

Notes: This table reports OLS estimates of baseline differences between control and treatment groups. For each variable, I report the baseline control group mean in column (1). In column (2), I report the estimated coefficient for the treatment indicator from a regression of the variable on the treatment indicator and stratification variables. In column (3), I report the *p*-value for the treatment indicator calculated using robust standard errors. I also report the RI *p*-value based on 5000 draws. In column (4), I report the sample size for the regression. †Senior managers at 19 control and 13 treatment factories reported not knowing the number of calls or were unaware of the helpline. *p<0.1; **p<0.05; ***p<0.01.

Table BII: Treatment effects: Truth-telling

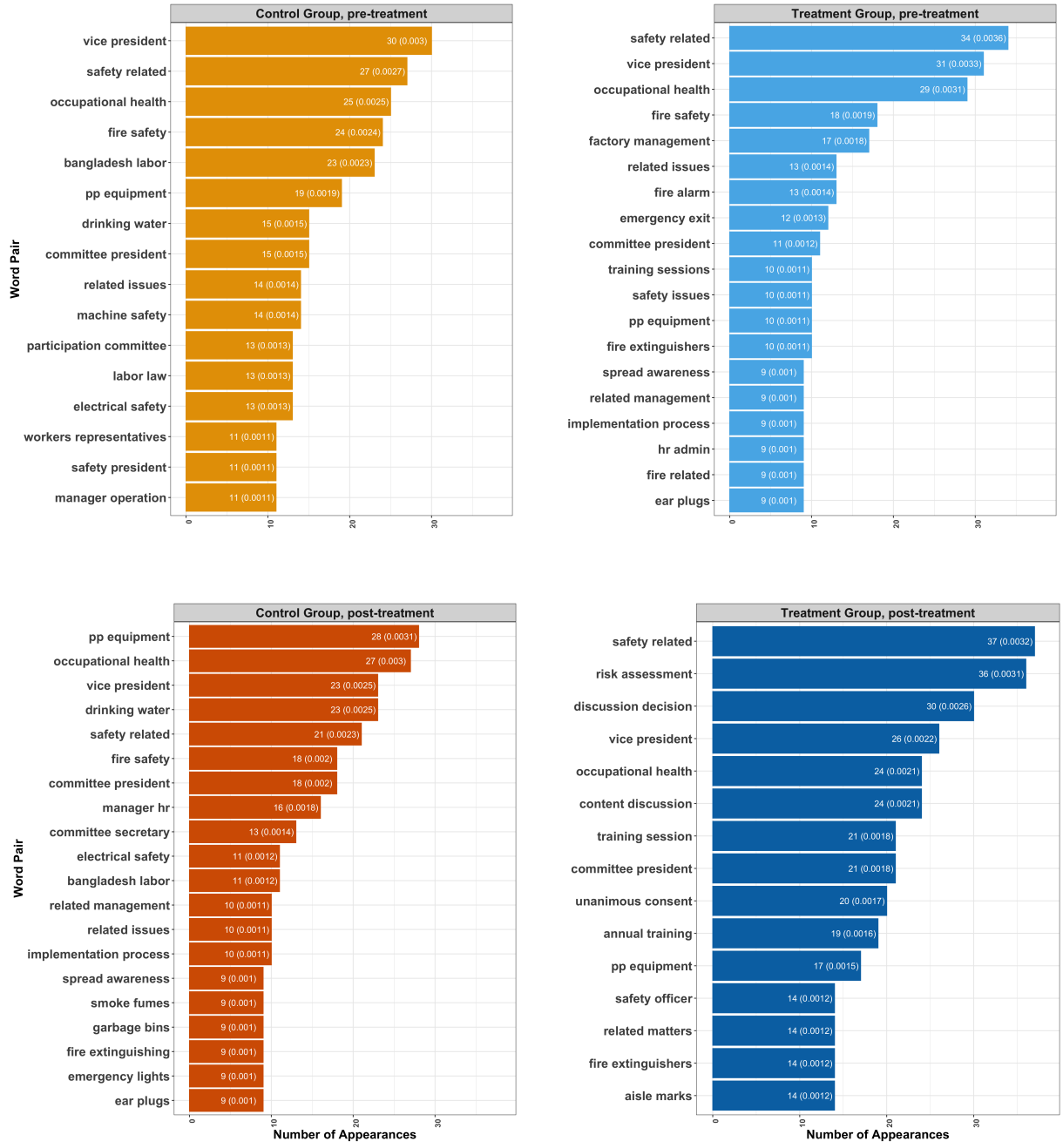
	Truth-telling		Awareness	
	Over-reports CAP completion	Under-reports helpline calls†	Correctly reports whether CAP visit	Aware of helpline
	(1)	(2)	(3)	(4)
Panel A: Senior Managers				
Treat	0.028	-0.149	0.251**	0.060
	(0.0742)	(0.1249)	(0.0990)	(0.0414)
	[0.675]	[0.226]	[0.015]	[0.042]
Control Mean	0.220	0.471	0.220	0.951
Observations	78	67	79	79
Strata FE	Y	Y	Y	Y
Control, base. dep. var.	Y	N	Y	Y

Panel B: Compliance with other Alliance Programs			
	Worker Helpline		
	Share workers with card	Share lower-level managers with card	Safety Training Record
	(1)	(2)	(3)
Treat	0.015	-0.065	0.023
	(0.0356)	(0.0515)	(0.0231)
	[0.701]	[0.217]	[1.000]
Control Mean	0.838	0.799	0.977
Observations	80	80	80
Strata FE	Y	Y	Y
Control, base. dep. var.	Y	Y	Y

Notes: This table reports OLS estimates of treatment effects on measures of truth-telling and of awareness. Outcome variables are listed at the top of each column. Robust standard errors are reported in round brackets. RI *p*-values based on 5000 draws are reported in square brackets. †Senior managers at 7 control and 5 treatment factories reported not knowing the number of calls or were unaware of the Alliance's worker helpline at the second data collection visit. *p<0.1; **p<0.05; ***p<0.01.

C: Figures and Tables

Figure CI: Most common two-word combinations in SC Meetings Minutes



Notes: To prepare the meeting minutes for text analysis, I strip the text of factory and participant names, the phrases "[health and] safety committee(s)" and "meeting(s)," English language stop words, numbers, and punctuation. I also replace the commonly used acronym of "ppe," which stands for personal protective equipment, and the complete phrase, with "pp equipment." Finally, I "stem" words, or replace them with their root, using the Porter stemmer. These approaches are common practice in text analysis (Gentzkow, Kelly and Taddy, 2019).

Figure CII: Correlation between WMS Management Index (excluding meeting question) and WMS Meeting-related Question, apparel firms in all countries

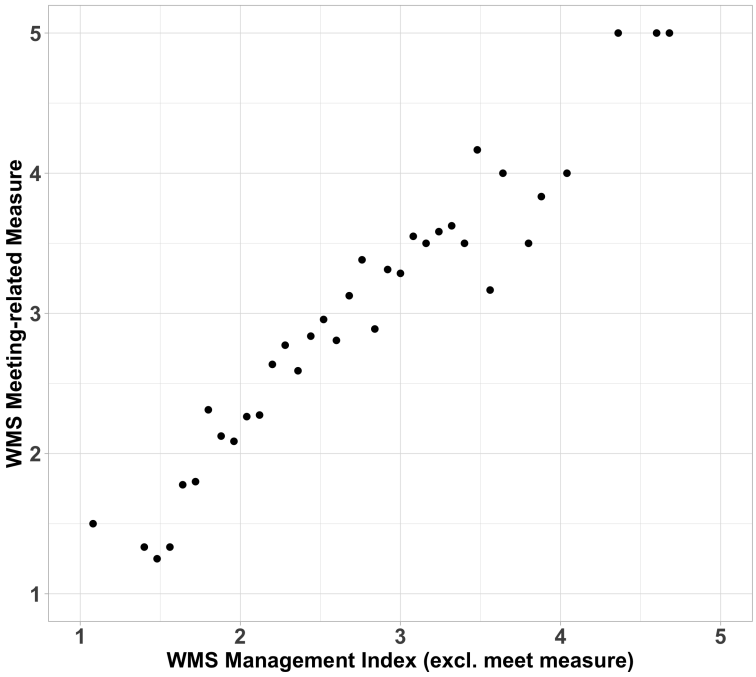


Table CI: Sample Summary Statistics

	(1) Mean	(2) SD	(3) Minimum	(4) Maximum	(5) N
Panel A: Primary outcome variables					
Compliance index	-0.037	0.409	-2.478	0.585	84
Safety Indicators index	-0.020	0.494	-1.226	1.157	84
Job satisfaction & well-being index	-0.056	0.461	-2.329	0.772	84
Number of employees	1124	1321	52	7633	80
Gross wages (log)	15.72	1.09	13.35	18.26	72
Labor productivity (log)	1.761	1.501	0.079	6.483	75
Panel B: Factory characteristics					
Trade union at factory (1=Yes)	0.024	0.153	0	1	84
EPZ(1=Yes)	0.179	0.385	0	1	84
Sewing (only)	0.405	0.494	0	1	84
Number product types	1.321	0.779	0	4	84
Monthly absenteeism (%)	4.588	3.845	0.074	26.916	80
Monthly turnover (%)	3.911	4.789	0.000	29.950	84
Prop. employees visit medical clinic (daily)	0.013	0.019	0.001	0.109	62
Participation in Alliance training (6 mo pre-baseline)	0.060	0.238	0	1	84
Number Alliance remediation visit to factory (6 mo pre-baseline)	0.179	0.415	0	2	84
Panel C: Worker survey respondent characteristics					
Age	27.51	3.42	21.55	40.29	84
Proportion female	0.50	0.28	0	1	84
Education (yrs)	6.53	1.62	2.75	11.3	84
Tenure (yrs)	3.80	2.30	0.43	14.38	84
Prior industry experience (yrs)	1.54	0.97	0	5.68	84
Panel D: SC President survey respondent characteristics					
Age	39.711	9.125	22	63	83
Female	0.108	0.313	0	1	83
Education (yrs)	15.867	1.999	8	18	83
Tenure (yrs)	7.152	6.227	0.083	25	83
Prior industry experience (yrs)	6.275	7.434	0	28.5	83
Panel E: SC Worker Representative survey respondent characteristics					
Age	27.169	5.098	19.5	48	83
Proportion Female	0.452	0.346	0	1	83
Education (yrs)	8.38	2.771	0	14	83
Tenure (yrs)	4.857	3.956	0.375	24.125	83
Prior industry experience (yrs)	1.728	1.969	0	8.5	83
Panel F: Senior Manager survey respondent characteristics					
Age	43.762	8.641	24	68	84
Female	0.024	0.153	0	1	84
Education (yrs)	15.929	1.974	8	18	84
Tenure (yrs)	8.9	8.405	0.083	42	84
Prior industry experience (yrs)	9.089	9.436	0	43	84

Notes: The sample size changes across rows due to differential data availability. In Panels D and E, the sample size is 83 factories because one factory was found not to have a true SC at baseline.

Table CII: Baseline balance tests, secondary outcome variables

	Control mean (SD)	T-C diff	p-value [RI p]	Number of factories
	(1)	(2)	(3)	(4)
Panel A: Secondary outcomes for workers, full sample				
Perceived SC compliance & effectiveness index	0.000 (0.5510)	-0.170	0.192 [0.187]	84
Perceived worker-manager relations index	0.000 (0.4065)	-0.187	0.141 [0.137]	84
Worker empowerment index	0.000 (0.4020)	-0.192	0.137 [0.131]	84
Worker organization awareness index	0.000 (0.7213)	-0.165	0.309 [0.315]	84
Number non-pecuniary benefits	6.483 (0.8817)	-0.404**	0.041 [0.035]	84
Monthly safety-related calls (per 1000 workers), Alliance Worker Helpline	0.057 (0.1446)	0.025	0.698 [0.875]	80
Monthly non-safety-related calls (per 1000 workers), Alliance Worker Helpline	0.422 (0.6764)	0.130	0.728 [0.935]	80
Panel B: Secondary outcomes for workers, dropping outlier on worker outcomes				
Perceived SC compliance & effectiveness index	0.000 (0.5510)	-0.126	0.312 [0.318]	83
Perceived worker-manager relations index	0.000 (0.4065)	-0.147	0.234 [0.212]	83
Worker empowerment index	0.000 (0.4020)	-0.117	0.272 [0.265]	83
Worker organization awareness index	0.000 (0.7213)	-0.121	0.442 [0.444]	83
Number non-pecuniary benefits	6.483 (0.8817)	-0.388*	0.051 [0.053]	83
Monthly safety-related calls (per 1000 workers), Alliance Worker Helpline	0.057 (0.1446)	0.025	0.699 [0.873]	79
Monthly non-safety-related calls (per 1000 workers), Alliance Worker Helpline	0.422 (0.6764)	0.144	0.703 [0.920]	79
Panel C: Secondary outcomes for factories				
Average Weekly Working Hours	54.37 (5.127)	2.34**	0.044 [0.051]	79
Efficiency (sewing section)	0.530 (0.1350)	0.070	0.203 [0.215]	33
Defects per hundred units	3.37 (3.097)	-1.01	0.133 [0.127]	72
Supplier-buyer relations index	0.018 (0.5671)	-0.133	0.411 [0.437]	71

Notes: This table reports OLS estimates of baseline differences between control and treatment groups. For each outcome, I report the baseline control group mean in column (1). In column (2), I report the estimated coefficient for the treatment indicator from a regression of the outcome or covariate on the treatment indicator and stratification variables. In column (3), I report the p-value for the treatment indicator calculated using robust standard errors. I also report the RI p-value for the coefficient reported in column (2) based on 5000 draws. In column (4), I report the sample size for the regression. *p$<$0.1; **p$<$0.05; ***p$<$0.01.

Table CV: Baseline balance tests, dropping outlier on worker outcomes

	(1) Control mean (SD)	(2) T-C diff	(3) p-value [RI p]	(4) N
SC Compliance index	0.000	-0.086	0.370	83
	(0.276)		[0.379]	
Safety Indicators	0.002	-0.032	0.770	83
	(0.401)		[0.765]	
Job satisfaction & well-being index	-0.005	-0.061	0.490	83
	(0.369)		[0.479]	
Number employees	1166	-242	0.360	84
	(1193)		[0.432]	
Gross wages (log)	15.820	-0.255	0.316	71
	(0.2529)		[0.303]	
Labor productivity (log)†	1.569	0.457	0.217	74
	(1.4240)		[0.201]	
Labor productivity (log)†, product FE	0.004	-0.028	0.858	74
	(0.6091)		[0.855]	

Notes: This table reports OLS estimates of baseline differences between control and treatment groups after dropping the outlier on worker outcomes. For each outcome, I report the baseline control group mean in column (1). In column (2), I report the estimated coefficient for the treatment indicator from a regression of the outcome on the treatment indicator and stratification variables. In column (3), I report the RI p-value for the coefficient reported in column (2) based on 5000 draws. In column (4), I report the sample size for the regression. †The regression sample is trimmed at the 1st and 99th percentile of all factory-month labor productivity observations. *p$<$0.1; **p$<$0.05; ***p$<$0.01.

Table CIII: Treatment effects: Secondary Outcome Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Worker Secondary Outcome Variables							
	Perceived SC compliance & effect index	Perceived worker-manager relations index	Worker empowerment index	Worker organization aware index	Number non-pecuniary benefits	Safety-related calls per 1k workers Alliance Helpline	Non-safety-related calls per 1k workers Alliance Helpline
Treatment effect	0.195	-0.039	0.067	0.016	-0.052	-0.002	-0.030
	(0.1216)	(0.0811)	(0.913)	(0.0793)	(0.1677)	(0.0753)	(0.1011)
	[0.094]	[0.630]	[0.487]	[0.842]	[0.761	[0.980]	[0.780]
Control mean	-0.109	0.072	-0.178	0.073	6.802	0.107	0.303
Observations	80	80	80	80	80	400	400
Factories	80	80	80	80	80	80	80
Stratification variables	Y	Y	Y	Y	Y	Y	Y
Control, base. dep. var.	Y	Y	Y	Y	Y	Y	Y
Panel B: Output & Factory Secondary Outcome Variables							
	Log(Output)	Average Weekly Working Hours	Efficiency (Sewing section)	Defects per 100 units	Supplier-buyer relations index		
Treatment effect	0.026	-0.550	0.018	0.275	0.047	0.015	
	(0.0885)	(0.5619)	(0.0267)	(0.2366)	(0.0544)	(0.0655)	
	[0.788]	[0.368]	[0.486]	[0.270]	[0.430]	[0.843]	
Factories	75	79	33	72	71	80	
Observations	374	395	165	360	355	400	
Stratification variables	Y	Y	Y	Y	Y	Y	
Control, baseline dep. var.	Y	Y	Y	Y	Y	Y	

Notes: This table reports OLS estimates of treatment effects on secondary outcome variables for workers and for factories. Each column in the table reports the estimated coefficient from a separate regression. The dependent variable in each column is regressed on the treatment indicator, stratification variables, and a control for the baseline value of the dependent variable. In Panel A, the regression sample changes in Columns (6) and (7) due to a different source of data for these outcomes. Each regression in Columns (1)-(5) includes one post-treatment observations per factory, while those in Columns (6)-(7) include five post-treatment observations per factory, where each observation is one month. In Columns (1)-(5), robust standard errors are reported in round brackets. In Panel A, Columns (6)-(7), and in Panel B, standard errors clustered at the factory level are reported in round brackets. In Panel B, for the supplier-buyer relations index, Column (4) includes all three variables from Table **??**, and Column (5) drops the third variable, which is missing for nine factories. Each regression includes five post-treatment observations per factory, where each observation is one month. RI *p*-values based on 5000 draws are reported in square brackets. For index variables, in all cases, higher values of the index correspond to "positive" outcomes. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table CIV: Pre-Analysis Plan Deviations

	Pre-Analysis Plan (PAP)	Modification
1)	The original PAP, which was posted on the AEA Registry in June 2017, included measures from survey question production line-level managers in the index of factory safety indicators (primary outcome variable). It also included survey measures from this group in certain secondary outcome index variables (perceived SC effectiveness and compliance index, worker-manager relations index, perception of worker capabilities index, and a perception of worker well-being index).	During baseline data collection, the research team determined that the line level managers were difficult to engage in surveys during the workday. As Safety Committees mostly aim to serve workers and to support senior management in occupational health and safety policies, after the baseline, I decided to remove variables from lower-level managers' survey data from these indices. I posted an updated PAP reflecting this change in December 2017.
2)	The SC Compliance Index included the following sub-variable: "President is management member and Vice President is worker member."	Due to an oversight, the information about the vice president's status was not collected for the first 57 baseline visits. I do not include this variable in the analysis.
3)	The Safety Indicators Index for the endline round included eight spotcheck variables that were only to be checked at endline.	Due to an administrative error, these additional items were not included in the checklist for the third visit for 14 out of 80 factories. As such, I do not include these variables in the analysis.
4)	The PAP indicated that I would also report outcomes measured at the individual-level at the individual level.	Consistent with best practice in econometric analysis for clusterd RCTs (Athey and Imbens, 2016), as the factory is the unit of analysis that is of interest, I omit individual-level regressions for space reasons.
5)	The PAP indicated that I would test for heterogeneous treatment effects by factories' location inside versus outside of an EPZ.	Ultimately, there were only 14 factories (7 treatment, 7 control) located in EPZs. There are large differences between these groups. As such, while I report baseline balance, I do not test for heterogeneous treatment effects by factory location.
6)	The PAP included hiring and machine downtime as factory-level secondary outcome variables.	I was not able to pilot the factory questionnaire until after I registered the PAP. I learned that many Alliance-covered factories did not systematically track these variables. Many factories in the sample had difficulty reporting them or indicated that they were unable to do. As such, I omit these secondary outcomes.
7)	The worker-manager relations index, a secondary outcome variable, include one variable to measure worker participation in strikes.	I decided that that participation in strikes was too sensitive to credibly measure in my setting, so I do not include the strike variable in the analysis.
8)	The PAP included an index of worker accidents and illness as a secondary outcome variable that included the medical clinic records, factory-reported illnesses and fires, and self-reported accidents and illness from workers.	The research team determined that factories' records of fires and accidents were often incomplete. For this reason, and due to the concern that the intervention may increase reporting of accidents by workers, I determined that the medical clinic records provide the most objective measure of accidents and illnesses.

Table CVI: Treatment effects: Primary outcomes, dropping outlier on worker outcomes

	(1)	(2)	(3)
Panel A	SC Compliance	Safety Indicators	Worker job satisfaction & Mental Well-being
Treatment effect	0.204***	0.139**	-0.013*
	(0.0593)	(0.0677)	(0.0793)
	{0.007}	{0.121}	{0.128}
	[0.001]	[0.052]	[0.075]
Control Mean	0.046	0.103	-0.013
Factories	79	79	79
Observations	79	79	79
Stratification variables	Y	Y	Y
Control, baseline dep. var.	Y	Y	Y
Panel B	Log(Labor productivity)		
Treatment effect	0.096	0.091	0.044
	(0.0661)	(0.0608)	(0.0410)
			{0.281}
	[0.129]	[0.169]	[0.300]
Factories	74	74	73
Observations	363	363	365
Stratification variables	Y	Y	Y
Control, baseline dep. var.	Y	Y	Y
Product type FE	N	Y	N
Trimmed sample	Y	Y	N
Dropping outlier	N	N	Y
Panel C	Log(Gross wages)	Log(Employment)	
Treatment effect	-0.015	-0.009	
	(0.0297)	(0.0215)	
	{0.521}	{0.521}	
	[0.647]	[0.718]	
Factories	71	79	
Observations	355	395	
Stratification variables	Y	Y	
Control, baseline dep. var.	Y	Y	

Notes: This table reports OLS estimates of treatment effects on primary outcome variables after dropping the outlier on worker outcomes. Each column in the table reports the estimated coefficient from a separate regression. The regression sample changes across Panels B and C due to differential data availability. Robust standard errors are reported in round brackets. RI p-values based on 5000 draws are reported in square brackets. p-values adjusted to control the False Discovery Rate (FDR) across primary outcomes are reported in curly brackets. *p<0.1; **p<0.05; ***p<0.01.

Table CVII: Baseline balance tests, SC President, SC Worker Representative, and Senior Manager Survey Respondents

	Control mean (SD)	T-C diff	p-value [RI p]	Number of factories
	(1)	(2)	(3)	(4)
Panel A: SC Presidents				
Age	40.857	-2.161	0.284	83
	(10.0912)		[0.273]	
Female	0.071	0.074	0.297	83
	(0.2607)		[0.314]	
Education (yrs)	16.119	-0.534	0.236	83
	(1.7137)		[0.242]	
Tenure (yrs)	6.313	1.552	0.261	83
	(5.5718)		[0.256]	
Prior industry experience (yrs)	7.347	-1.858	0.235	83
	(8.5988)		[0.246]	
Panel B: SC Worker Representatives				
Age	26.964	0.403	0.711	83
	(4.368)		[0.716]	
Proportion Female	0.488	-0.069	0.375	83
	(0.340)		[0.433]	
Education (yrs)	8.452	-0.073	0.902	83
	(2.572)		[0.907]	
Tenure (yrs)	4.476	0.738	0.385	83
	(4.023)		[0.393]	
Prior industry experience (yrs)	2.05	-0.700	0.104	83
	(2.063)		[0.108]	
Panel C: Senior Managers				
Age	43.791	-0.152	0.935	84
	9.6005		[0.940]	
Female	0.023	-0.000	1.000	84
	(0.1525)		[1.000]	
Education (yrs)	16	-0.131	0.755	84
	(1.7995)		[0.820]	
Tenure (yrs)	9.535	-1.317	0.462	84
	(8.9144)		[0.457]	
Prior industry experience (yrs)	8.481	1.228	0.552	84
	(10.2333)		[0.569]	

Notes: This table reports OLS estimates of baseline differences between control and treatment survey participant groups. For each covariate, I report the baseline control group mean in column (1). In column (2), I report the estimated coefficient for the treatment indicator from a regression of the covariate on the treatment indicator and stratification variables. In column (3), I report the p-value for the treatment indicator calculated using robust standard errors. I also report the RI p-value for the coefficient reported in column (2) based on 5000 draws. In column (4), I report the sample size for the regression. In Panels A and B, the sample size is 83 factories because one factory was found not to have a true SC at baseline. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table CVIII: Local Average Treatment Effects (LATEs): Primary Outcome Variables

	(1)	(2)	(3)
Panel A	SC Compliance	Safety Indicators	Job Satisfaction & Mental Well-being
LATE	0.220	0.154	-0.161
	(0.059)***	(0.069)**	(0.079)**
Control mean	0.046	0.103	-0.013
Observations	80	80	80
Stratification variables	Y	Y	Y
Control, base. dep. var.	Y	Y	Y
Panel B	Log(Labor Productivity)	Log(Gross Wages)	Log(Employment)
LATE	0.041	-0.017	-0.011
	(0.039)	(0.031)	(0.023)
Observations	370	360	400
Factories	74	72	80
Stratification variables	Y	Y	Y
Control, base. dep. var.	Y	Y	Y
Dropping outlier	Y	N	N

Notes: This table reports two stage least squares (2SLS) estimates of treatment effects on primary outcome variables. Each column in the table reports the estimated coefficient from a separate regression. In Panel A, higher values of index variables correspond to more "positive" outcomes. Robust standard errors are reported in parentheses. In Panel B, the regression sample changes across columns due to differential data availability. For labor productivity, results are shown after dropping the control factory that partially shuts down during the study. Each regression includes five post-treatment observations per factory, where each observation is one month. Standard errors clustered at the factory level are reported in round brackets. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table CIX: Lee (2009) bounds for primary outcome index variables

	Lower bound	Upper bound
SC compliance index	0.205	0.217
	(0.076)***	(0.069)***
Safety indicators index	0.140	0.141
	(0.091)	(0.072)**
Job satisfaction & mental well-being index	-0.159	-0.158
	(0.081)**	(0.084)*

Notes: This table reports Lee treatment effect bounds for sample selection. Outcome variables are listed on the left. Column (1) reports the lower bound. Column (2) reports the upper bound. Standard errors are reported in parentheses. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table CX: Baseline balance tests, cheap & costly compliance & sub-index components

	(1) Control mean	(2) T-C diff	(3) p-value [RI p]	(4) N
Panel A: SC Compliance				
Formation sub-index	0.000	-0.171	0.303	84
	(0.5332)		[0.304]	
Operations sub-index	0.000	0.035	0.788	84
	(0.5440)		[0.781]	
Responsibilities sub-index	0.000	-0.096	0.366	84
	(0.4167)		[0.347]	
Panel B: Safety Indicators				
CAP completion sub-variable	0.017	0.092	0.661	84
	(0.9940)		[0.655]	
Worker awareness sub-index	0.000	-0.547**	0.028	84
	(0.8917)		[0.027]	
Worker knowledge sub-index	0.000	-0.082	0.665	84
	(0.8191)		[0.661]	
Senior manager awareness sub-variable	0.000	0.348	0.130	84
	(1.000)		[0.136]	
Panel C: Worker Job Satisfaction and Mental Well-being				
Job satisfaction sub-index	0.000	-0.179	0.262	84
	(0.7407)		[0.248]	
Mental well-being sub-index	0.000	-0.237	0.220	84
	(0.5571)		[0.213]	
Turnover sub-variable	0.000	-0.002	0.991	84
	(1.000)		[0.989]	
Absenteeism sub-variable	-0.027	0.150	0.386	84
	(0.9835)		[0.408]	

Notes: This table reports OLS estimates of baseline differences between control and treatment groups for the sub-indexes and sub-variables that comprise each primary outcome index. Each panel reports the sub-index/sub-variable balance tests for a different outcome variable. For each sub-index or sub-variable, column (1) reports the baseline control group mean. Column (2) reports the estimated coefficient for the treatment indicator from a regression of the sub-index or sub-variable on the treatment indicator and stratification variables. Column (3) reports the RI p-value for the coefficient reported in column (2) based on 5000 draws. The regression sample remains the same in all rows. *p<0.1; **p<0.05; ***p<0.01.

Table CXI: Treatment effects: Meeting Minutes & Visitors to Medical Clinic

	(1)	(2)	(3)
	Meeting Minutes Log(Word Count)	Mean proportion of workforce visits medical clinic (daily)	
Panel A: Main treatment effects			
Treatment effect	0.227	-0.0017*	-0.0019
	(0.1372)	(0.0010)	(0.0012)
	[0.094]	[0.107]	[0.135]
Factories	71	62	62
Observations	71	254	254
Control mean		0.011	0.012
Stratification variables	Y	Y	Y
Control, baseline dep. var.	Y	Y	Y
Weighted regression	N	Y	N
Panel B: Heterogeneous treatment effects by managerial practices			
Below Median	0.185	-0.0012	-0.0016
	(0.1972)	(0.0013)	(0.0017)
	[0.375]	[0.422]	[0.464]
Above Median	0.278	-0.0024	-0.0022
	(0.2797)	(0.0015)	(0.0014)
	[0.142]	[0.200]	[0.222]
p-val, diff	0.740	0.519	0.771
	[0.745]	[0.626]	[0.834]
Factories	71	62	62
Observations	71	254	254
Control mean, below median		0.010	0.014
Control mean, above median		0.011	0.011
Stratification variables	Y	Y	Y
Control, baseline dep. var.	Y	Y	Y
Weighted regression	N	Y	N

Notes: This table reports OLS estimates of treatment effects on the number of words in Meeting Minutes for SC meetings and the proportion of employees that visit the medical clinic on a daily basis. Each column in the table reports the estimated coefficient from a separate regression. In column (2), probability weights based on the number of pre-treatment observations are applied. The dependent variable in each column is regressed on the treatment indicator, stratification variables, and a control for the baseline value of the dependent variable. Standard errors clustered at the factory level are reported in round brackets. RI *p*-values based on 5000 draws are reported in square brackets. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table CXII: Treatment effects: Worker awareness & Workforce composition

Panel A: Worker awareness	(1)	(2)	(3)	(4)	(5)	(6)
	Aware of SC & its responsibilities	Knows factory has SC	Knows how to report safety concern to SC	Reported num SC resp	Reports SC as channel for raising issues	Knows SC members
Treatment	0.053**	0.040**	0.011	-0.118	0.063	0.073**
	(0.0249)	(0.0177)	(0.0232)	(0.1228)	(0.0379)	(0.0380)
	[0.036]	[0.017]	[0.651]	[0.345]	[0.101]	[0.028]
Control mean	0.843	0.945	0.920	3.060	0.655	0.689
Observations	80	80	80	80	80	80
Stratification variables	Y	Y	Y	Y	Y	Y
Control, baseline dep. var.	Y	Y	Y	N	Y	N
Panel B: Workforce composition						
	Age	Female	Tenure	Prior exp.	Yrs. Education	
Treatment	−0.171	−0.041	0.251	0.041	0.264	
	(0.5122)	(0.0368)	(0.3204)	(0.1833)	(0.2669)	
	[0.741]	[0.274]	[0.437]	[0.828]	[0.336]	
Control mean	27.667	0.577	3.696	1.507	6.635	
Observations	80	80	80	80	80	
Stratification variables	Y	Y	Y	Y	Y	
Control, baseline dep. var.	Y	Y	Y	Y	Y	

Notes: This table reports OLS estimates of treatment effects on worker SC awareness variables and on workforce characteristics. In Panel A, the first four columns report outcomes included in the Safety Indicators Index (prior to standardization for inclusion in the index). Each column in the table reports the estimated coefficient from a separate regression. The regression sample is the same in all columns. The dependent variable in each column is regressed on the treatment indicator, stratification variables, and a control for the baseline value of the dependent variable (if available). RI p-values based on 5000 draws are reported in square brackets. *p<0.1; **p<0.05; ***p<0.01.

Table CXIII: Treatment effects: Physical indicators of factory safety

	Control mean	ITT Effect	RI p-value
	(1)	(2)	(3)
Factory safety spotcheck index	0.000	0.217**	[0.015]
		(0.0892)	
Sewing: Machines have guards *and* workers wear PPE† for their task	0.500	0.076	[0.619]
		(0.1512)	
Cutting: Machines have knife guards *and* workers wear PPE for their task	0.792	0.071	[0.557]
		(0.1173)	
Dyeing and jobs handling chemicals: Safety masks, goggles, gloves, aprons, and boots worn by workers handling chemicals	0.545	0.102	[0.668]
		(0.2293)	
All PPE appropriate size, functional, and well-maintained	0.951	0.050	[0.492]
		(0.0350)	
Aisles clearly marked and markings visible	0.780	0.025	[1.000]
		(0.0908)	
Aisles clear of sewing scrapes and debris	0.951	0.048	[0.503]
		(0.0338)	
Aisles clear of obstruction	0.854	0.014	[1.000]
		(0.0800)	
Machines in good working order & dangerous parts properly covered	0.927	0.070	[0.247]
		(0.0404)	
Work stations maintained in tidy condition (no loose materials close to electrical appliances)	0.976	0.022	[1.000]
		(0.0228)	
One or more easily accessible first aid kit in section	0.976	0.022	[1.000]
		(0.0228)	
Physical separation between storage & production areas	0.976	0.022	[1.000]
		(0.0228)	
Drinking water easily accessible for all workers	1.000	-0.025	[1.000]
		(0.0252)	
Drinking water provided appears clean (visual check)	1.000	-0.025	[1.000]
		(0.0252)	
Stratification variables		Y	

Notes: This table reports OLS estimates of treatment effects on the spotcheck sub-index and for each variable in the spotcheck index. Four variables on the spotcheck checklist drop from the analysis because all factories were found to comply with these variables (see Table **??**). Sub-variables are listed on the left. Results are shown for the sub-variables *prior* to standardizing them for inclusion in the index. Column (1) reports the control group mean of the outcome variable. Column (2) reports the estimated ITT effect from a regression of the outcome variable on the treatment indicator and stratification variables. Robust standard errors are reported in round brackets. Column (3) reports RI p-values based on 5000 draws are reported in square brackets. †PPE stands for personal protective equipment. PPE vary by task and include equipment such as eye guards, finger guards, chain mesh gloves, goggles, boots, etc. *p<0.1; **p<0.05; ***p<0.01.

Table CXIV: Treatment effects: Workers' job satisfaction & mental well-being sub-variables

	Control mean	ITT Effect
	(1)	(2)
Panel A: Job Satisfaction		
Self-reported job satisfaction (qualitative scale, coded 1-5)	4.813	-0.045 (0.0486) [0.384]
Respondent suggested/helped family or friends to get a job at their factory (previous 4 months)	0.573	-0.049 (0.0428) [0.266]
Respondent has thought about leaving their job at factory for safety-related reasons (previous 3 months)	0.024	0.019* (0.0101) [0.064]
Panel B: Mental Well-being		
Self-reported level of stress in life (qualitative scale, coded (-1)-(-5))	-1.760	-0.059 (0.0755) [0.474]
Self-reported perceived extent of control over their life (qualitative scale, coded 1-5)	4.082	-0.035 (0.0557) [0.534]
Self-reported perceived extent of control safety at factory (qualitative scale, coded 1-5)	4.368	-0.037 (0.0584) [0.530]
Self-reported stress about experiencing accident or injury at factory (qualitative scale, coded (-1)-(-5))	-1.488	0.039 (0.0599) [0.538]
Self-reported frequency of feeling unsafe at factory (qualitative scale, coded (-1)-(-5))	-1.236	-0.013 (0.0317) [0.691]
Panel C: Turnover and Absenteeism		
Turnover	3.356	0.051 (0.3107) [0.884]
Absenteeism	4.457	0.3866 (0.2507) [0.162]
Observations		80
Stratification variables		Y
Control, base. dep. var.		Y

Notes: This table reports OLS estimates of treatment effects on each variable included in the worker job satisfaction and mental well-being index. Each panel reports the sub-variable results for a different sub-index. Sub-indexes and sub-variables are listed on the left. Results are shown for the variables *prior* to orienting them to be unidirectional and standardizing them for inclusion in the index. Column (1) reports the control group mean of the outcome variable. Column (2) reports the estimated ITT effect from a regression of the outcome variable on the treatment indicator, stratification variables, and a control for the baseline value of the outcome variable. Robust standard errors are reported in round brackets. RI p-values based on 5000 draws are reported in square brackets. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table CXV: Ex post minimum detectable effect sizes (MDEs): Business competitiveness outcomes

	Control mean (sd)	MDE
	(1)	(2)
Log(Labor productivity)†	1.477	0.169
	(1.345)	
Log(Labor productivity), dropping labor productivity outlier	1.455	0.102
	(1.344)	
Log(Gross wages)	15.865	0.081
	(1.080)	
Log(Employment)	6.665	0.060
	(1.038)	

Notes: This table reports ex post power calculations and minimum detectable effect sizes for labor productivity, employment, and wage outcome variables with 80% power at the 5% significance level. Outcome variables are listed on the left. Column (1) reports the control group mean and standard deviation in column. Column (2) reports the ex post MDE. †Reported MDE is for sample trimmed at the 1st and 99th percentiles of all factory-month observations for labor productivity.

Table CXVI: Baseline balance tests within subgroups for management practices heterogeneity analysis, primary outcome variables

	(1) Control mean	(2) T-C diff	(3) p-value [RI p]	(4) N
Below median subgroup:				
SC Compliance	0.025	0.113	0.228	40
	(0.3035)		[0.233]	
Cheap SC compliance	0.014	0.106	0.393	40
	(0.4632)		[0.376]	
Costly SC compliance	0.038	0.123	0.258	40
	(0.3055)		[0.273]	
Safety Indicators	0.099	-0.067	0.667	40
	(0.4801)		[0.651]	
Job Satisfaction & Mental Well-being	0.070	-0.198	0.254	40
	(0.4036)		[0.284]	
Log(Labor productivity)†, trimmed sample	0.0450	-0.236	0.371	40
	(0.6025)		[0.315]	
Log(Wages)	15.625	0.007	0.983	38
	(1.1540)		[0.984]	
Log(Employment)	6.297	0.060	0.861	40
	(1.0979)		[0.858]	
Above median subgroup:				
SC Compliance	-0.032	-0.114	0.228	40
	(0.2602)		[0.255]	
Cheap SC compliance	-0.016	-0.109	0.431	40
	(0.3837)		[0.412]	
Costly SC compliance	-0.052	-0.120	0.386	40
	(0.3971)		[0.384]	
Safety Indicators	-0.067	0.013	0.933	40
	(0.3399)		[0.934]	
Job Satisfaction & Mental Well-being	-0.053	-0.021	0.827	40
	(0.3505)		[0.838]	
Log(Labor productivity)†, trimmed sample	-0.0153	-0.087	0.748	35
	(0.6272)		[0.731]	
Log(Wages)	16.004	-0.376	0.372	34
	(0.9429)		[0.372]	
Log(Employment)	6.925	-0.514	0.190	40
	(0.8741)		[0.166]	

Notes: This table reports OLS estimates of baseline differences between control and treatment groups within above- and below-median management subgroups for treatment effect heterogeneity analysis. For each outcome, I report the baseline control group mean in column (1). In column (2), I report the estimated coefficient for the treatment indicator from a regression of the outcome or covariate on the treatment indicator and stratification variables. In column (3), I report the p-value for the treatment indicator calculated using robust standard errors. I also report the RI p-value for the coefficient reported in column (2) based on 5000 draws. Column (4) reports the number of observations in that subgroup. † The regression also includes product-type fixed effects. The trimmed sample drops factory-month observations in the 1st and 99th percentiles of labor productivity. *p<0.1; **p<0.05; ***p<0.01.

Table CXVII: Heterogeneous treatment effects by managerial practices, cheap and costly compliance and business competitiveness outcomes

	Cheap Compliance	Costly Compliance	Log(Labor productivity)		Log(Gross wages)	Log(Employment)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Treatment Phase						
Below median	0.114	0.118	0.161	0.049	-0.015	-0.010
	(0.0789)	(0.1480)	(0.1020)	(0.0466)	(0.0365)	(0.0308)
	[0.169]	[0.449]	[0.127]	[0.403]	[0.673]	[0.754]
Above Median	0.017	0.653***	0.017	0.030	-0.018	-0.007
	(0.0873)	(0.1079)	(0.0732)	(0.0588)	(0.0484)	(0.0323)
	[0.841]	[0.000]	[0.830]	[0.662]	[0.731]	[0.858]
p-val, diff	0.400	0.005	0.278	0.803	0.965	0.940
	[0.419]	[0.016]	[0.281]	[0.828]	[0.968]	[0.941]
Factories	80	80	75	74	72	80
Observations	80	80	368	370	360	400
Panel B: Post-treatment phase						
Below median	0.089	0.196	0.005	0.029	-0.004	-0.009
	(0.1108)	(0.1734)	(0.0632)	(0.0539)	(0.0418)	(0.0384)
	[0.435]	[0.288]	[0.957]	[0.712]	[0.929]	[0.850]
Above Median	0.078	0.538***	-0.027	-0.038	-0.015	0.023
	(0.1076)	(0.1629)	(0.0543)	(0.0523)	(0.0522)	(0.0477)
	[0.525]	[0.005]	[0.636]	[0.504]	[0.808]	[0.671]
p-val, diff	0.940	0.142	0.690	0.356	0.865	0.615
	[0.944]	[0.215]	[0.758]	[0.500]	[0.877]	[0.624]
Factories	80	80	75	74	72	80
Observations	80	80	218	222	216	240
Panel C: Pooled						
Below median	0.101	0.157	0.101	0.042	-0.011	-0.010
	(0.0673)	(0.1174)	(0.0617)	(0.0423)	(0.0361)	(0.0328)
	[0.168]	[0.254]	[0.166]	[0.499]	[0.759]	[0.801]
Above Median	0.048	0.596***	0.001	0.004	-0.017	0.004
	(0.0690)	(0.0964)	(0.0589)	(0.0507)	(0.0420)	(0.0368)
	[0.547]	[0.000]	[0.987]	[0.939]	[0.727]	[0.922]
p-val, diff	0.567	0.004	0.246	0.572	0.915	0.782
	[0.616]	[0.037]	[0.279]	[0.661]	[0.928]	[0.784]
Factories	80	80	75	74	72	80
Observations	160	160	586	592	576	640
Stratification variables	Y	Y	Y	Y	Y	Y
Control, baseline dep. var.	Y	Y	Y	Y	Y	Y
Product type FE	N	N	Y	Y	N	N
Trimmed sample	N	N	Y	N	N	N
Dropping outlier	N	N	N	Y	N	N

Note: This table reports OLS estimates of heterogeneous treatment effects on cheap and costly compliance and on business competitiveness variables. Each outcome variable is indicated at the top of the table. For labor productivity, in column (3), the sample is trimmed at the 1st and 99th percentile of all factory-month observations. In column (4), a factory in the control that partially shut down during the study is dropped. Each panel reports the results for data collection during the treatment phase, after the treatment phase, and pooling both rounds of data collection. In each panel, the "Below median" row reports the estimated treatment effect for the subgroup with below median baseline managerial practices. In each panel, the "Above median" row reports the estimated treatment effect for the subgroup with above median baseline managerial practices. The final two rows in each panel report the *p*-value of the difference between the estimated treatment effects for below and above median subgroups. RI *p*-values based on 5000 draws are reported in square brackets. All regressions include stratification variables and a control for the baseline value of the dependent variable. Standard errors clustered at the factory level are reported in round brackets. *p<0.1; **p<0.05; ***p<0.01.

Table CXVIII: Treatment effects after end of intensive enforcement: Business competitiveness outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A	Log(Labor productivity)					
Treat	-0.028	-0.017	-0.009	-0.022	-0.010	0.003
	(0.0475)	(0.0438)	(0.0483)	(0.0443)	(0.0393)	(0.0424)
				{1.000}		{1.000}
	[0.555]	[0.737]	[0.702]	[0.642]	[0.838]	[0.957]
Factories	75	75	75	74	74	74
Observations	218	218	218	222	222	222
Stratification variables	Y	Y	Y	Y	Y	Y
Control, baseline dep. var.	Y	Y	N	Y	Y	N
Product FE	N	Y	N	N	Y	N
PDS Lasso Selected Controls	N	N	Y	N	N	Y
Dropping outlier	N	N	N	Y	Y	Y
Panel B	Log(Gross wages)		Log(Employment)			
Treat	-0.009	-0.011	0.003	0.005		
	(0.0313)	(0.0341)	(0.0292)	(0.0293)		
	{1.000}	{1.000}	{1.000}	{1.000}		
	[0.807]	[0.761]	[0.926]	[0.867]		
Factories	72	72	80	80		
Observations	216	216	240	240		
Stratification variables	Y	Y	Y	Y		
Control, baseline dep. var.	Y	N	Y	N		
PDS Lasso Selected Controls	N	Y	N	Y		

Notes: This table reports OLS estimates of the persistence of treatment effects on labor productivity, employment, and gross wages measured 3-4 months after the end of the intensive enforcement period. Outcome variables are listed at the top of each column. Each column reports the estimated ITT effect from a separate regression. Panel A reports results for labor productivity. In columns (1)-(3), the sample is trimmed at the 1st and 99th percentile of all factory-month labor productivity observations. In columns (4)-(6), a factory in the control that partially shut down during the study is dropped. In Panel B, each regression includes five post-treatment observations per factory, where each observation is one month. The regression sample changes across columns due to differential data availability. Standard errors clustered at the factory level are reported in round brackets. *p*-values adjusted to control the FDR across primary outcomes are reported in curly brackets. RI *p*-values based on 5000 draws are reported in square brackets. In Panel A, Column (3), the PDS lasso selected control variables include the baseline value of the dependent variable, its square, an indicator for having a trade union, and an indicator for packaging production. In Column (6), the PDS lasso selected control variables include the baseline value of the dependent variable, an indicator for having a trade union, an indicator for packaging production, and an indicator for weaving production. In Panel B, Column (2), the PDS lasso selected control variables include the baseline value of the dependent variable, its square, an indicator for dyeing production, an indicator for packaging production, an indicator for weaving production, the mean log baseline number of employees, and the square of mean baseline absenteeism. In Column (4), they include the baseline value of the dependent variable, its square, an indicator for packaging production, and an indicator for weaving production. *p<0.1; **p<0.05; ***p<0.01.

D: Robustness checks for HTE analysis

I report robustness checks for the HTE analysis by management practices (Section 5.3). First, there is correlation in factories' characteristics: Better-managed factories tend to be somewhat larger and less compliant. This raise the concern that only one of these characteristics determines the intervention's effects. To examine this possibility, I regress each outcome on the treatment indicator, an indicator for each dimension of heterogeneity, and interactions between each dimension and the treatment. This specification demands a lot of the data, but it provides qualitative insight into the relative importance of each dimension. Table DI presents the results. For all three primary outcome index variables, management practices remain important. For the safety indicators index, while the interaction term loses statistical significance, it is largest in magnitude (col (2), p=0.139).

Another concern is that MNCs may more intensively monitor less compliant factories and that this generates the heterogeneous effects. In this case, one would expect the Alliance to be more likely to audit factories that, at baseline, are less compliant with the SC law. The Alliance audited five treatment factories during the study period, but all of the audits occurred after the 4-5 month data collection visit. As such, differential auditing could not drive the heterogeneous effect patterns in Panel A of Table VI.

Finally, I use an alternative measure of management practices. This measure captures a different dimension of managerial capacity: HR management. I measure HR practices using an index of worker-reported HR practices and relations with managers that I prespecified as a secondary outcome variable (see the Supplementary Materials for index components). I find a qualitatively similar pattern of heterogeneous effects using this variable as with my main measure. See Tables DII and DIII below.

Table DI: Testing the importance of each dimension of heterogeneity

	Compliance	Safety Indicators	Job satisfaction & mental well-being
	(1)	(2)	(3)
Treat	0.194	0.002	-0.572
	(0.1310)	(0.1475)	(0.2231)
	[0.225]	[0.990]	[0.013]**
Treat*Abv med Mgmt	0.235**	0.207	0.326*
	(0.1168)	(0.1378)	(0.1712)
	[0.087]	[0.155]	[0.068]
Treat*Abv med Size	-0.174	0.139	0.127
	(0.1137)	(0.1548)	(0.1494)
	[0.187]	[0.346]	[0.438]
Treat*Abv med Compliance	-0.040	-0.075	0.407**
	(0.1140)	(0.1546)	(0.1877)
	[0.766]	[0.621]	[0.040]
Observations	80	80	80
Stratification variables	Y	Y	Y

Notes: This table reports OLS estimates of heterogeneous treatment effects, controlling for all dimensions of heterogeneity. Each column in table the reports the estimated coefficients from a separate regression. The regression sample is the same in all columns in a panel. Robust standard errors are reported in round brackets. RI p-values based on 5000 draws are reported in square brackets. *p<0.1; **p<0.05; ***p<0.01.

Table DII: Baseline balance tests within non-management subgroups for heterogeneity analysis, primary outcome variables

	Control mean (SD)	T-C diff	*p*-value [RI *p*]	Number of factories
	(1)	(2)	(3)	(4)
Panel A: Factory Size				
Below median subgroup:				
SC Compliance	-0.003	0.078	0.437	40
	(0.2793)		[0.442]	
Safety Indicators	0.021	0.083	0.616	40
	(0.3869)		[0.627]	
Job Satisfaction & Mental Well-being	0.004	0.026	0.868	40
	(0.4419)		[0.868]	
Above median subgroup:				
SC Compliance	-0.010	-0.075	0.567	40
	(0.2831)		[0.536]	
Safety Indicators	-0.007	-0.118	0.456	40
	(0.4384)		[0.464]	
Job Satisfaction & Mental Well-being	-0.002	-0.133	0.367	40
	(0.3170)		[0.452]	
Panel B: SC Compliance				
Below median subgroup:				
SC Compliance	-0.224	-0.061	0.439	40
	(0.1961)		[0.369]	
Safety Indicators	-0.029	-0.059	0.689	40
	(0.3238)		[0.688]	
Job Satisfaction & Mental Well-being	-0.063	0.030	0.804	40
	(0.4146)		[0.808]	
Above median subgroup:				
SC Compliance	0.221	0.027	0.502	40
	(0.1241)		[0.543]	
Safety Indicators	0.042	0.162	0.289	40
	(0.4914)		[0.314]	
Job Satisfaction & Mental Well-being	0.068	-0.161	0.320	40
	(0.3255)		[0.360]	
Panel C: Location in EPZ				
EPZ subgroup:				
SC Compliance	-0.190	0.385	0.182	14
	(0.3620)		[0.188]	
Safety Indicators	-0.051	0.220	0.571	14
	(0.3606)		[0.560]	
Job Satisfaction & Mental Well-being	-0.104	0.503*	0.076	14
	(0.4362)		[0.078]	
Non-EPZ subgroup:				
SC Compliance	0.031	-0.013	0.868	66
	(0.2473)		[0.866]	
Safety Indicators	0.017	-0.066	0.597	66
	(0.4239)		[0.584]	
Job Satisfaction & Mental Well-being	0.022	-0.173	0.151	66
	(0.3648)		[0.137]	
Panel D: HR Managerial Practices				
Below median subgroup:				
SC Compliance	-0.051	0.080	0.452	40
	(0.3239)		[0.476]	
Safety Indicators	-0.181	0.045	0.806	40
	(0.4545)		[0.798]	
Job Satisfaction & Mental Well-being	-0.011	-0.250	0.121	40
	(0.3929)		[0.151]	
Above median subgroup:				
SC Compliance	0.031	-0.095	0.376	40
	(0.2322)		[0.347]	
Safety Indicators	0.167	-0.055	0.674	40
	(0.2917)		[0.680]	
Job Satisfaction & Mental Well-being	0.011	0.040	0.710	40
	(0.3679)		[]	

Notes: This table reports OLS estimates of baseline differences between control and treatment groups within non-management subgroups for treatment effect heterogeneity analysis. In Panels A, B, and D, I partition the sample into above/below median subgroups using the baseline value of the variable. For location in Export Processing Zone (EPZ), I partition the sample using this variable. Each panel reports the within subgroup baseline differences for a different dimension of heterogeneity. For each outcome, I report the baseline control group mean in column (1). In column (2), I report the estimated coefficient for the treatment indicator from a regression of the outcome on the treatment indicator and stratification variables. In column (3), I report the *p*-value for the treatment indicator calculated using robust standard errors. I also report the RI *p*-value for the coefficient reported in column (2) based on 5000 draws. In column (4), I report the sample size for the regression. *p<0.1; **p<0.05; ***p<0.01.

Table DIII: Other heterogeneous treatment effects: Primary Outcome Index Variables, Pooling treatment and post-treatment rounds of data

	SC Compliance	Safety Indicators	Job Satisfaction & Mental Well-being
	(1)	(2)	(3)
Panel A: Baseline Size			
Below median	0.234***	0.083	-0.114
	(0.0743)	(.0767)	(0.0970)
	[0.004]	[0.279]	[0.283]
Above median	0.163*	0.109	0.084
	(0.0824)	(0.0878)	(0.0930)
	[0.060]	[0.236]	[0.314]
p-val, diff	0.516	0.834	0.167
	[0.575]	[0.831]	[0.131]
Panel B: Baseline SC Compliance			
Below median	0.236**	0.098	-0.045
	(0.0967)	(0.0909)	(0.1001)
	[0.029]	[0.304]	[0.678]
Above median	0.180***	0.099	0.007
	(0.0612)	(0.0737)	(0.0827)
	[0.010]	[0.202]	[0.929]
p-val, diff	0.625	0.991	0.687
	[0.676]	[0.990]	[0.705]
Panel C: Baseline HR Management Practices			
Below median	0.129	0.024	-0.096
	(0.0827)	(0.0685)	(0.1062)
	[0.121]	[0.754]	[0.388]
Above median	0.288***	0.181**	0.064
	(0.0825)	(0.0843)	(0.0746)
	[0.002]	[0.056]	[0.388]
p-val, diff	0.187	0.148	0.231
	[0.220]	[0.208]	[0.221]
Observations	160	160	160
Factories	80	80	80
Stratification variables	Y	Y	Y
Control, base. dep. var.	Y	Y	Y

Note: This table reports OLS estimates of heterogeneous treatment effects on primary outcome index variables, pooling treatment and post-treatment rounds of data. Each outcome variable is indicated at the top of the table. Each panel reports the results for a different dimension of heterogeneity. In each panel, the "Below median" row reports the estimated treatment effect for the subgroup with below median baseline values of the heterogeneity variable. In each panel, the "Above median" row reports the estimated treatment effect for the subgroup with above median baseline values of the heterogeneity variable. The final row in each panel reports the *p*-value of the difference between the estimated treatment effects for below and above median subgroups. All regressions include stratification variables and a control for the baseline value of the dependent variable. All subgroups have 40 factories. Robust standard errors are reported in round brackets. RI *p*-values based on 5000 draws are reported in square brackets. Index variables constructed using Anderson (2008) variance-covariance weighted index. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.