

## OFFICE OF TRADE AND LABOR AFFAIRS CONTRACT RESEARCH PROGRAM

U.S. Department of Labor  
Bureau of International Labor Affairs (ILAB)

Download this and other papers at  
<http://www.dol.gov/ilab/reports/>

The opinions expressed are those of the author(s). They are offered for the purpose of discussion and contribution to research. They are not statements of official Department of Labor or ILAB policy.

## FOREWORD

We contracted with Dr. Andrew Schrank to produce this paper, "Improving Labor Inspections Systems: Design Options." We conceived of this as a methodological-research exercise, with the hopes of empirical execution in the future.

We thank Dr. Schrank for his excellent work and view this paper as a first step toward a longer-term goal to aid labor ministries in the design of cost-effective and successful inspection systems. The use in Dr. Schrank's analysis of examples from Latin America reflects a likely scope of future engagement (through cooperative mechanisms defined in free trade agreements) and a desire to narrow the scope of consideration so as to keep the analysis specific and concrete.

Future implementation of a method(s) identified by this contract is subject to considerable uncertainty; including, the many technical factors identified so ably by Dr. Schrank, and importantly, the identification of appropriate and interested partner countries. Should ILAB have the future capability and resources to pursue a method(s), it will be with the full cooperation of an interested partner country, and only after fulsome and formal diplomatic engagement.

Kenneth A. Swinnerton  
Chief, Economic and Labor Research  
Office of Trade and Labor Affairs  
September 30, 2014

**Improving Labor Inspections Systems:  
Design Options**

United States Department of Labor  
Bureau of International Labor Affairs  
Purchase Order: DOLB139K34960

**Author**  
Andrew Schrank

May 2014

**Submitted to:**  
Sarah Gormly Donovan  
Contracting Officer's Representative  
US Department of Labor  
Bureau of International Labor Affairs  
Email: [donovan.sarah@dol.gov](mailto:donovan.sarah@dol.gov)  
Phone: (202) 693-4872

**Submitted by:**  
Andrew Schrank  
Professor of Political Science  
University of New Mexico  
Albuquerque, NM 87131  
Email: [schrank@unm.edu](mailto:schrank@unm.edu)  
Phone: (203) 809-2458

## **Introduction.**

Labor inspectors have been portrayed as “foot soldiers of the campaign for decent work” (Piore and Schrank 2008, p. 21) around the world. They visit firms, farms, and factories, determine whether—and to what degree—they are meeting their obligations to their legal and contractual obligations to their workers, and issue warnings, sanctions, and/or advice in the event of noncompliance (Richthofen 2002, pp. 100-3). Examples would include the “compliance officers” who work for the Occupational Safety and Health Administration and the Wage and Hour Division of the United States Department of Labor, the “work environment inspectors” who monitor safety, health, and hours of work in Sweden, and the “generalist” labor inspectors who oversee most of the labor and employment laws in France and the Iberian world.

Do inspectors really improve compliance and, if so, by what means do they achieve their goals? The answers are anything but obvious, for neither inspectors nor their strategies—broadly defined to include everything from who is eligible to become an inspector in the first place to how they are trained and treated by their supervisors to the specific practices they use in particular cases—are randomly distributed across countries and communities, and experts therefore have trouble distinguishing the effects of inspection itself from the background factors that determine when, where, how, and by whom inspections occur. When analysts have trouble distinguishing the effects of “treatments” like public policies and practices from the effects of covariates or confounding factors that influence their distribution across time and space, however, they increasingly use experimental methods to isolate the effects of the former from the effects of the latter, and randomized trials are widely considered to be the “gold standard” among experimental methods (Barrett and Carter 2010, p. 516).<sup>1</sup> “By randomly assigning treatment,” explain Stephen Fienberg and his associates, “we eliminate the influence of any possible confounder on the putative cause, and thus create a situation in which the resulting raw association is the appropriate measure of the causal effect” (Fienberg et al. 2003, p. 38).

The following paper identifies experimental designs for the evaluation of labor inspection systems in Latin America. It includes six principal sections. Section 1 discusses the main differences between the “Latin model” (Piore and Schrank 2008) of labor inspection and the more familiar approach adopted by enforcement agencies like OSHA and the Wage and Hour Division in the US. Section 2 discusses theories of regulatory noncompliance and develops a logic model that links enforcement strategies to compliance outcomes in the region. Section 3 discusses some of the strategies that are available to Latin American labor inspectors and sets the stage for a discussion of their assignment to experimental subjects. Section 4 identifies five possible subjects of experimentation (e.g., inspectors, firms, jurisdictions) and discusses their respective receptivity to both random assignment and counterfactual analysis (e.g., data needs, estimation procedures, etc.). Section 5 addresses practical considerations involved in the design and conduct of experiments on inspection systems—including their utility, ethics, and viability—and introduces a checklist designed to facilitate their assessment. And Section 6 describes three potential experiments—labeled “professionals v. partisans,” “risk-based targeting v. randomized

---

<sup>1</sup> See, however, Ravallion (2009) and Deaton (2010) for a more skeptical take on randomized trials and Barrett and Carter (2010) for a seemingly balanced treatment.

inspection,” and “carrots v. sticks” respectively—and discusses their principal goals and limitations in light of the checklist.

Several caveats are necessary at the outset. First, by “enforcement strategies” I mean not only strategies that are prone to experimental manipulation but strategies that are immune to manipulation for legal, logistical, or ethical reasons. I refer to the former as “plausible” strategies and the latter as “theoretical” strategies and focus on the former for contractual reasons. Second, by “compliance outcomes” I mean levels of compliance with prevailing legal standards regardless of their perceived merits. I take no position on the proper scope or substance of Latin American labor and employment law. Third, by “inspectors” I mean the officials responsible for “most aspects of working conditions and health and safety” (Vega Ruiz 2009, p. 15) in the region including—but not limited to—wages and hours. Latin American labor inspectors have advisory as well as supervisory authority and are therefore more autonomous and less specialized than their North American counterparts (Piore and Schrank 2008; Bensusán 2009; Casale 2012; Bhorat and Stanwix 2013). And, finally, by taking the roots of noncompliance seriously, I hope to identify treatments that allow the region’s inspectors to make the most of their autonomy and authority. The latter goal is particularly important, for I not only share Robert Kagan and John Scholz’s sense that “reliance on any single theory of noncompliance is likely to be wrong” (Kagan and Scholz 1984, p. 85) in principle but agree that “when translated into an enforcement strategy, it is likely to be counterproductive” in practice as well. Consider, for example, a theory that traces noncompliance to avarice, and thus calls for strict sanctions that render noncompliance more costly than compliance, when the real problem is inefficiency, and the employer’s *inability* to compete and comply simultaneously. While capacity-building measures (e.g., on-site technical advice, workforce development, etc.) might bring the inefficient employer into compliance over time, sanctions would be more likely to drive him underground or out of business—a particularly vexing problem in a labor-abundant developing country.

### **1. Models of labor inspection.**

The boundaries of the proposed experimental interventions are necessarily limited by the rights and responsibilities of the labor inspectors themselves. What are their rights and responsibilities? Wolfgang von Richthofen of the International Labour Office (ILO) describes a traditional distinction between “generalist” and “specialist” approaches. “The former describes systems where inspectors have a broad mandate (usually under a comprehensive labour code) to deal with matters related to a host of functions,” he explains, like occupational safety and health, wages and hours, industrial relations, and/or vocational education and training. The latter, by way of contrast, describes systems that deal “in the main with only one of these major labour inspection functions—usually occupational safety, health and the working environment” (Von Richthofen 2002, p. 37).

But Richthofen goes on to describe a second-order distinction rooted less in jurisdictional considerations than in strategic concerns. “Thus, while sanctioning models are concerned mainly with punishable contraventions or violations of rules and regulations, compliance systems secure conformity with the law (and beyond), without necessarily using formal methods of enforcement” including the imposition of fines and penalties that nonetheless “remain available as a last resort” (Von Richthofen 2002, p. 37).

Labor inspection systems can thus be classified in terms of their divisions of labor and their dependence on sanctions. Figure 1 distills and illustrates the logical possibilities.

**Figure 1: A typology of labor inspection systems**

		Division of labor	
		<i>Higher/specialized</i>	<i>Lower/generalist</i>
Dependence on sanctions	<i>Higher/sanctioning</i>	The United States	Several Eastern European countries (e.g., Hungary)
	<i>Lower/compliance</i>	Scandinavia	France, Spain, Portugal, and most of their former colonies including Brazil, Chile, the Dominican Republic, Mexico, Ecuador, Uruguay

The United States tends to divide responsibility for labor law enforcement among number of different agencies (e.g., OSHA, the Wage and Hour Division, the Equal Employment Opportunity Commission, the National Labor Relations Board, and their state and local analogs, etc.) that tend to respond to noncompliance by issuing sanctions, and the US approach has therefore been described as “both a specialized and a sanctioning system” (Schrank and Piore 2007, p. 10). By way of contrast, Latin American inspectors inherited the “generalist” model found in France and the Iberian Peninsula and have the right “to decide whether to impose sanctions, give advice or request remedial action” (Casale 2012, p. 13) in the event of noncompliance. “In this context,” explains Giuseppe Casale, “most countries in the region have labour inspectors who wield general power over social issues (some countries in the region refer to a ‘comprehensive inspection’), encompassing most aspects of working conditions and health and safety such as Argentina, Brazil, Chile, Uruguay and/or other matters, such as cooperatives in Argentina” (2012, p. 13). Other countries combine a more narrow division of labor with a “compliance model,” in the case of Scandinavia, or a broad division of labor with a preference for sanctions, in the case of Eastern Europe. For instance, Danish inspectors are exclusively responsible for “occupational safety and health issues” (EPSU 2012, p. 22) but have a broad range of tools at their disposal, whereas Hungarian inspectors have responsibility for almost every aspect of labor and employment law but lack “the discretionary power to decide whether or not to impose a fine” (EPSU 2012, pp. 12-6) in the event of noncompliance. If they discover a violation of Hungarian labor law, they have no choice but to sanction the employer.

In short, Latin American labor inspectors have broad rights and responsibilities that offer ample scope for experimentation. The question, therefore, is how to design an experiment that allows their supervisors to make the most of their rights and responsibilities, and the answer necessarily depends on the hypothesized link between enforcement strategies and compliance outcomes.

## **2. Logic model**

Our goal is to understand the relationship between enforcement strategies and compliance outcomes. But diagnosis necessarily comes before treatment. A medical intervention is unlikely to succeed if the doctor misdiagnoses the patient, and labor law enforcement is no different. An enforcement strategy that is designed with the wrong diagnosis in mind

is unlikely to foster compliance—and may well prove counterproductive. And the logic model must therefore take diagnosis seriously.

What are the causes of noncompliance? Mainstream economic theories that take rational utility maximization as their starting points assume that regulations impose costs on firms—otherwise they would desist from the proscribed activities on their own—and that firms are owned and operated by “amoral calculators” (Kagan and Scholz 1984, p. 68) who will evade regulations, and their attendant costs, whenever profitable. On the mainstream account, therefore, the inspector’s job is make noncompliance more burdensome and less rational than compliance not only by raising the cost of detection and punishment *ex post* but by raising the probability of detection and punishment *ex ante*. “The goal,” explain Kagan and Scholz, “is deterrence. The governmental inspector, accordingly, should be a strict *policeman*, indifferent to businessmen’s manipulative excuses” (1984, p. 68).

The mainstream account offers a valuable point of departure, for firms *are* in business to make a profit and *will* tend to cut costs, and perhaps even break the law, when profitable. But mainstream approaches have trouble accounting for firms that comply with costly obligations when detection and punishment are unlikely (or unimpressive) or breach their obligations when detection and punishment are common and compelling, and a rival body of literature therefore adopts a thicker portrait of human motivation and a more nuanced model of organizational decision making. While the former acknowledges the importance of altruism, solidarity, status, and social norms among individuals, the latter allows for the possibility of “external contingencies, political battles, unacknowledged cultural beliefs, and formal and informal internal pathologies that undercut both the determination of goals and their achievement” (Vaughan 1998, p. 29) among firms. And proponents of the alternative account therefore treat compliance as the rule—rather than the exception to be explained—and trace noncompliance not simply to avarice but to ignorance—including not only ignorance of the law but ignorance of organizational or productive techniques that would allow the firm to compete and comply simultaneously.

The point is less to privilege a single account of noncompliance than to acknowledge and understand their differences, for the diagnosis adopted by a regulatory agency, whether implicitly or explicitly, not only “has real effects on enforcement practices” (Kagan and Scholz 1984, p. 69) but has real effects on regulatory outcomes as well. Consider, for example, an inspector who applies a mainstream diagnosis to an employer who is more ignorant than amoral. By imposing sanctions in lieu of education and training, he may add to her costs, undercut her profits, and drive her underground or out of business.

But efforts to educate or train amoral calculators can also backfire, and Kagan and Scholz therefore worry about bad decisions made in good faith by inspectors who are simply naïve or gullible. “If he is up against a truly ‘bad apple,’” they argue (1984, p. 84), “the consultant-inspector can be duped into not penalizing intentional violations.”

The more salutary outcomes are found in the northwestern and southeastern quadrants of Figure 2. On the one hand, amoral calculators should respond to the threat of punishment, and efforts to maximize the likelihood and cost of sanctions should therefore deter their malfeasance *ex ante* or bring them into compliance *ex post*. On the other hand, ignorant firms should respond to capacity building, and efforts to disseminate legal and productive

information should therefore allow them to reconcile compliance with competitiveness. (Of course, in practice an employer could be both ignorant and immoral, and a packaged strategy may therefore prove necessary.)

**Figure 2: A stylized model of enforcement strategies and outcomes in different organizational contexts**

		Context: cause of noncompliance	
		<i>Avarice</i>	<i>Ignorance</i>
Strategy: response to noncompliance	<i>Punishment</i>	Compliance via cost-benefit calculation and/or deterrence	Noncompliance (e.g., evasion and/or exit by ignorant enterprises)
	<i>Education</i>	Noncompliance (e.g., deceit or double-dealing by amoral enterprises)	Compliance via capacity building

The stylized model in Figure 2 implies that compliance outcomes are influenced not only by “enforcement strategies” defined narrowly to include “inspection strategies, penalty schemes, employer education/assistance, etc.” (ILAB 2013, p. 4) but by a broader range of *actors* than inspectors and *factors* than inspections including prosecutors who bring suit, judges who issue sanctions, mediators who provide non-adversarial alternatives, and financial and training institutions that offer compliance assistance.

Consider, for example, the question—or *questions*—of sanctions. Who is (or should be) responsible for their administration? While sanctions could be imposed by the inspectors themselves in theory, and would thus constitute a particularly “big stick,” they are often reserved for judges in practice, for legislators worry that corrupt inspectors would use their big sticks to extract bribes rather than compliance.<sup>2</sup> Are judges really less prone to extortion than inspectors? The answer purportedly depends on the personnel policies—e.g., recruitment, remuneration, and retention criteria—that govern the two professions in the different countries. How big should sanctions be? While bigger sanctions are a bigger deterrent, they are also a bigger stick, and thus constitute both a bigger threat to employer solvency (and jobs) and a greater source of temptation to judges as well as inspectors. Could they be replaced by less punitive remedies like compliance assistance measures that might help employers come into compliance without sacrificing competitiveness or mediation services that might resolve conflicts before they escalate and grow costly? The answers depend in part on the sources of noncompliance—i.e., ignorance or avarice—and in part on the availability of the relevant services, and they therefore tend to vary across countries, communities, and sectors.<sup>3</sup>

In other words, the range of enforcement strategies relevant to compliance *in theory* is all but limitless, and would include not only inspection strategies and tactics but judicial and bureaucratic structures and practices, the nature, level, and responsibility for sanctions, and industrial policies designed to “reconcile compliance and competitiveness” (Schrank

<sup>2</sup> For instance, Colombian inspectors are allowed to impose monetary penalties but in the Dominican Republic the authority to impose sanctions is reserved for the judiciary and is not available to the inspectors themselves (Casale 2012, p. 22).

<sup>3</sup> See Vega Ruiz and Robert (2013) for a more thorough discussion of sanctions and their administration.

2013, p. 306). But the range of enforcement strategies “available to labor inspectorates” (ILAB 2013, p. 4) *in practice* is necessarily limited by their jurisdictions, budgets, and histories, and the stylized enforcement model in Figure 2 thus implies a logic model that aligns treatment and diagnosis by treating workplace inspection as a multistage process that is circumscribed by: organizational inputs that are hard to manipulate in the short run (e.g., inspectors, information technology, vehicles, etc.); inspection strategies that are more amenable to manipulation in the short run (e.g., targeting, remediation, etc.); and contextual factors that are all but immune from manipulation in the short or even medium run (e.g., private sector productivity, public sector probity, the spatial distribution of employment).<sup>4</sup>

Consider, for example, the inspectors who constitute the lifeblood of the enforcement process. Skilled inspectors are not available on a spot market to be recruited at low cost by experimental researchers, and if they were recruited temporarily for the purposes of experimentation their incentive structures would be decidedly atypical in any event. After all, the existing literature implies that tenure guarantees are indispensable to public sector probity and performance (Rauch and Evans 2000; Schrank 2009), but tenure guarantees would be neither practical nor credible when offered to experimental subjects.

The character and concentration of private sector employment are neither less important nor more susceptible to experimental manipulation, and the logic model thus concentrates on inspection strategies that are subject to manipulation and ignores inputs and contextual factors for the time being:

Stage 1 is the *deployment* of investigators to firms or worksites and is constrained by the availability of inspectors, their access to transportation, and the spatial concentration or dispersal of employment in particular. Are inspectors available? Do they have access to vehicles? Are their targets geographically concentrated or dispersed? The answers will necessarily influence the quality as well as the number of inspections undertaken.

Stage 2 involves an *assessment* of wages and working conditions and is typically carried out by an individual inspector or team. Is the employer compliant and, if not, which laws or aspects of the labor code are being violated? Violations are typically classified by degree of severity (e.g., minor, serious, very serious) and prioritized accordingly (Ugarte Cataldo 2008, p. 190; Godínez Vargas 2011, pp. 7-9).

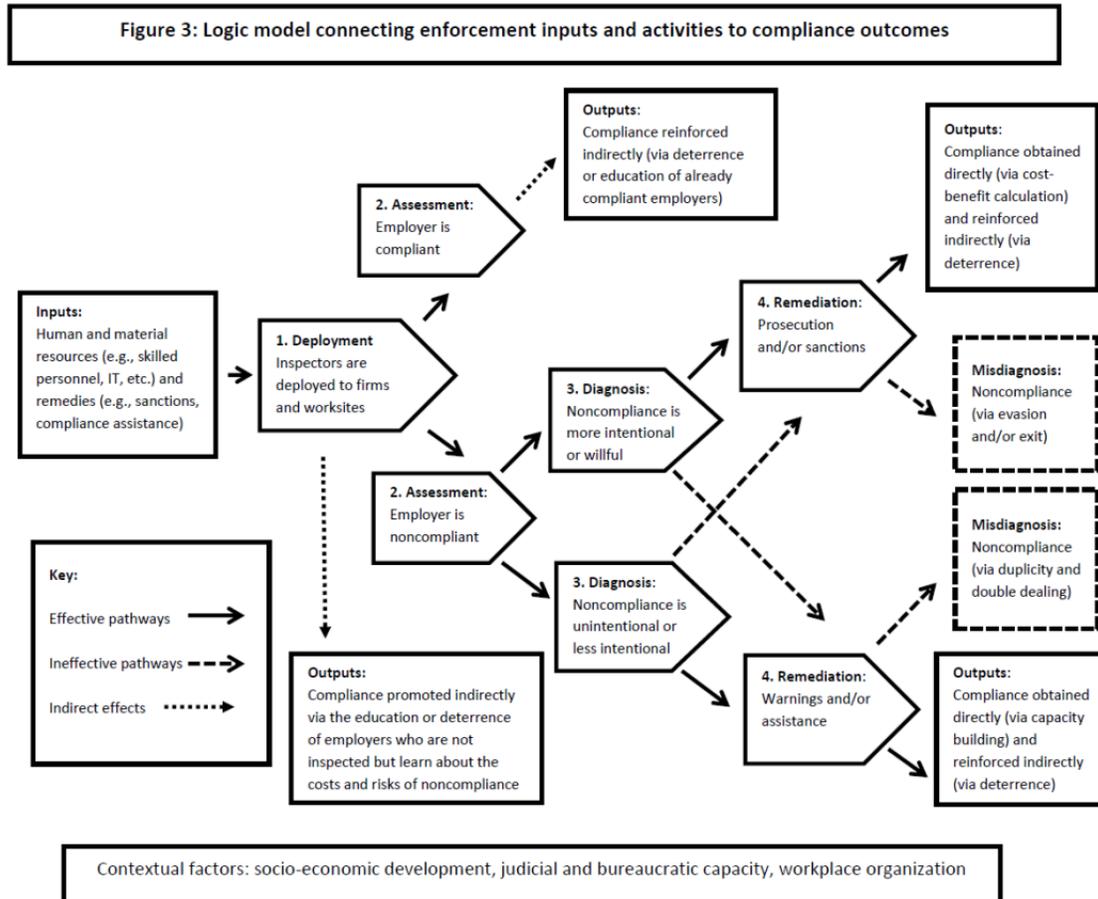
Stage 3 is *diagnostic* and presupposes the availability of *skilled* inspectors in particular. Is noncompliance more intentional and willful or unintentional and accidental? The answer is likely to fall along a continuum—from more to less deliberate—and is particularly important in light of the discretion available to Latin American inspectors and their dual—i.e., advisory and supervisory—role.

Stage 4 is *remedial* and entails the deployment of warnings, sanctions, and/or advice and support of various types in an effort to map the treatment (or remedy) to the diagnosis. In

---

<sup>4</sup> See Barrett and Carter (2010) on the impediments to experimentation on macroeconomic, political economy, and infrastructural questions, in general (p. 527), and contractual arrangements, in particular (p. 522).

theory, the availability of multiple remedies allows Latin American labor inspectors to address noncompliance in at least three different ways: first, by invoking sanctions (or the threat of sanctions) in an effort to raise the cost of transgression to amoral calculators; second, by providing (or facilitating access to) capacity-building measures in an effort to limit the need for transgression among more feeble firms; and, third, by putting ‘boots on the ground’ in an effort to disseminate knowledge and deter transgression more generally. Figure 3 offers a diagrammatic representation of the model as a whole.

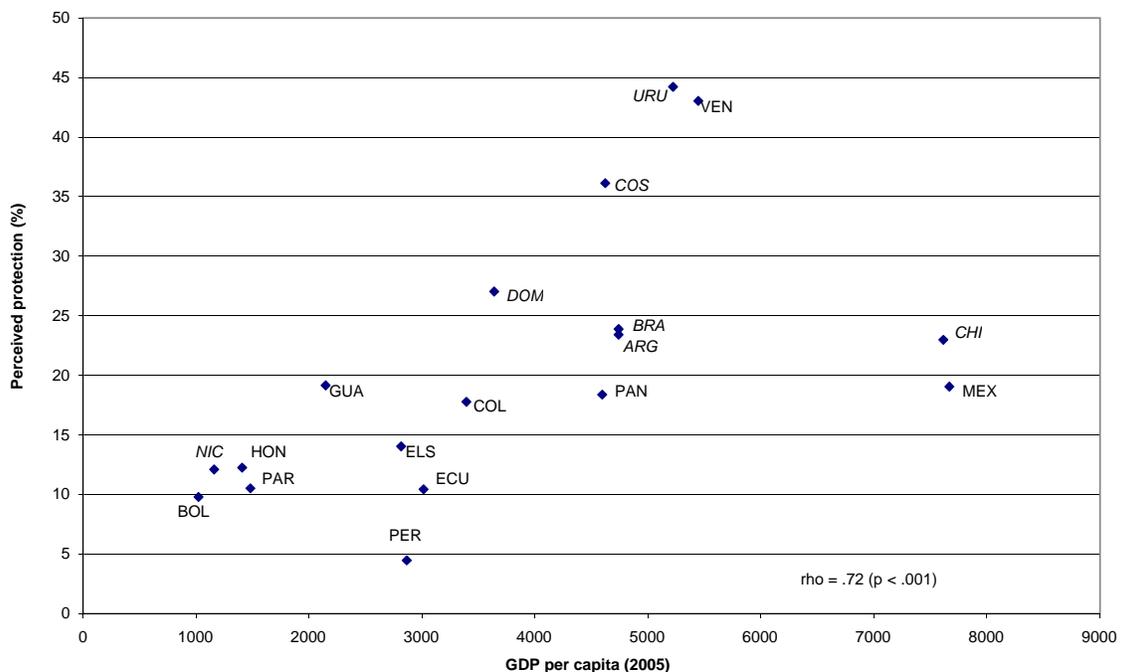


In short, the model assumes that inspections vary in terms of quality as well as quantity, and that their impact will therefore depend on both the human and material resources (e.g., skilled personnel, IT, vehicles, etc.) and legal remedies (e.g., sanctions, compliance assistance, etc.) available to the enforcement authorities. When the authorities have more human and material resources at their disposal, after all, they can commission or conduct more and better inspections, including planned investigations as well as post hoc visits; and when the inspectors themselves have access to an array of remedies, they can tailor their efforts to the situation at hand. But the broad responsibilities and vast discretion that allow the inspectors to tailor their enforcement efforts to the needs of particular firms also open the door to misdiagnosis and malpractice. When amoral calculators are treated with kid gloves, for example, or feeble firms are put through the proverbial wringer, the results

are likely to include duplicity and double dealing, on the one hand, or evasion and exit, on the other. In neither case is compliance likely, let alone sustainable.<sup>5</sup>

Unfortunately, however, the risk of misdiagnosis and malpractice are particularly high in Latin America, for the region tends to combine unproductive employers who are in dire need of capacity building with unskilled inspectors who are ill-equipped to meet their needs. Figure 4 underscores the depth of the problem by: first, plotting the percentage of salaried workers that feels “sufficiently” or “very” protected by labor law—an admittedly crude proxy for compliance—against gross domestic product (GDP) per capita in United States dollars—a similarly crude proxy for productivity—for all 18 non-communist Latin American countries in the year 2005; and, second, italicizing the abbreviations of the countries that appoint “properly trained inspectors” to career positions in a functioning civil service on the basis of “strict selection criteria” (Vega Ruiz 2009, p. 25).

**Figure 4: Perceived compliance and economic development in Latin America**



Data: Latinobarometer (2005); World Bank (2013); Vega Ruiz (2009). A regression of the natural log of perceived compliance on the natural log of GDP per capita and a dummy variable for civil service protection yields the following equation:  $Y = -1.13 + .48 \ln GDP_{pc} + .41 \text{CivServ} + e$  with an  $R^2$  of .49 and both predictors significant at  $p < .10$ .

The data imply that vulnerability is not only the regional norm but is felt most acutely in countries that lack both productive firms and professional inspectors, and they therefore have at least three implications. First, they underscore both the necessity and the context-

<sup>5</sup> Policymakers who want to tailor their efforts to the particularities of the case without leaving diagnosis up to the individual inspector have been known to use firm size as a crude proxy for capacity and have thus limited compliance assistance and related programs to small and mid-sized employers. For instance, Chile allows small firms to substitute training for fines on their first offense, and an initial impact evaluation revealed “slightly higher levels of compliance among training participants than among a control group that paid their fine” (ILO 2006, p. 18). By way of contrast, Tandler (2001) holds that size-based targeting gives small, informal firms an incentive to stay small and informal.

dependence of improved enforcement. While the Andean and Central American countries are dominated by less productive employers, and would thus do well to consider training their inspectors to “serve as *consultants* more often than as policemen” (Kagan and Scholz 1984, pp. 82-3), they are also confronted by corruption and cronyism that impose a “limit on consultative enforcement practices” (Kagan and Scholz 1984, p. 84). Second, they more or less limit the range of candidates for experimental analysis to the italicized countries. After all, the candidate countries should have relatively well functioning inspectorates; otherwise, the risk of noncompliance among the inspectors themselves will be heightened.<sup>6</sup> And third, they highlight the need for reforms that are either ill-suited to experimentation, beyond the reach of inspectors, or both. For example, the countries that are not italicized in Figure 3 would do well to reconsider their personnel policies in an effort to recruit and retain more professional inspectors (Piore and Schrank 2008). But personnel policies are frequently established by elected officials (who view public sector employment as a source of political patronage) or dictated by civil service commissions (that are bound by constitutional or legislative statute) and are therefore difficult—if not necessarily impossible—to randomize on an experimental basis.<sup>7</sup> The countries that are near the origin would do well to consider “productive development policies” (Melo and Rodríguez-Clare 2006) designed to bring their productivity levels up to the demands of their labor standards. But productive development policies are typically administered by planning ministries and development banks and are thus beyond the reach of most labor ministries. And Chile and Mexico would do well to reconsider their minimum wages, for the latter dictate the sanctions available to their inspectors as well as the wages available to their workers—by means of an indexing system that links the size of the sanction to the effective minimum wage and the severity of the violation (see, e.g., Melis 2006; Castillo 2013)—and have nonetheless “been held under control” (Bizberg 2012, p. 25) in recent years. In fact, Graciela Bensusán holds that the sanctions available to Mexican inspectors have been all but undermined by the collapse of the country’s minimum wage (Bensusán 2008, p. 104), and Hugo Guzmán notes that Chilean sanctions are “almost laughable” (Guzmán 2011; my translation) as well. But minimum wages are subject to neither the authority of the labor ministry nor experimental manipulation and would thus seem to fall outside of the scope of the project.

Corruption poses a particularly vexing problem in developing countries and thus warrants further discussion. After all, the logic model in Figure 3 will break down in the event of

---

<sup>6</sup> Results obtained in relatively well functioning inspectorates with capable human resources need not translate to their less fortunate or efficient counterparts, of course, but inspectorates that are at serious risk of corruption would be vulnerable to the type of experimental noncompliance that would compromise the very “integrity of data collection” (Barrett and Carter 2010, p. 523) in the first place—and would almost certainly be better off addressing their more fundamental human resource deficits before reconsidering their inspection strategies in any event.

<sup>7</sup> The debate over the Indian police experiments is illustrative. While Banerjee et al. (2012) found that in-service training and a freeze on transfers had a positive impact on perceived police performance in the Indian state, their critics conclude that their most interesting finding “was perhaps not so much what did or did not work when implemented but what was impossible to implement, even as an experiment. In spite of the presence of the team of an Indian Police Service officer and the full co-operation of the head of the police,” explains Lant Pritchett, “the station chiefs just did not implement key parts of the reform—in spite of producing ‘administrative facts’ claiming they were being implemented” (Pritchett 2013, p. 29; see also Moehler 2010, p. 39; Ravallion 2012, p. 111; Davis and Mota Prado 2014, p. 216; as well as Barrett and Carter 2010, p. 522, on the impediments to experimenting on “new contract terms” and the like).

widespread fraud and duplicity on the part of the inspectors themselves, but developing countries frequently lack the resources they would need to recruit and retain honest inspectors and to monitor the inspectors they do recruit. Efforts to improve enforcement in developing countries have therefore endeavored to combat corruption on the cheap by deploying inspectors in teams or rotating their areas of responsibility.<sup>8</sup> Others have endeavored to “monitor the monitors” (Banerjee et al. 2012; Duflo et al. 2013). But this paper concerns experimental evaluations of strategies that might “affect firm compliance with labor law and regulation” and are at least potentially available to inspectors, and that assumes engagement with a reasonably functional inspectorate in the first place. I therefore do not address anti-corruption strategies. But I do caution ILAB to focus its experimentation efforts on countries with seemingly low corruption and to build methods for collecting at least qualitative data on corruption into their research design to augment contextual analysis (e.g., anonymous “I paid a bribe” surveys of firms, focus groups etc.).

In short, the range of theoretically interesting strategies is far broader than the range of plausible experimental interventions. And I therefore focus on *plausible* strategies in the pages to follow.

### **3. Enforcement strategies.**

By *plausible* strategies, I mean strategies that are under the control of the labor ministry, responsible for the process of labor inspection, and subject to experimental manipulation. The administrators who oversee the inspectors typically have some degree of control over their recruitment, remuneration, retention, evaluation, training, deployment, methods, *and/or* responses to noncompliance, and by exercising their control in five broadly defined areas—personnel policies, intake procedures, target selection, field methods, and remediation—they develop their enforcement strategies. I will briefly discuss each of the five areas in turn.

*Personnel policies.* By “personnel policies” I mean anything that shapes the identities, skills, or treatment of the inspectors themselves. Examples would include policies that determine who is eligible to be an inspector in the first place (e.g., academic credentials, degree requirements, gender quotas, etc.); how they are chosen from among the eligible candidates (e.g., by means of competitive exams, personal connections, interviews, etc.); how they are evaluated (e.g., performance metrics) and compensated (e.g., fixed salaries, pay-for-performance, etc.) once they are on the job; when and how they are trained (e.g., scope, content, and timing); and the terms of their contracts (e.g., fixed term, open-ended, civil service protections, etc.). The logical combinations are endless and are believed to influence the quality of the human resources dedicated to the enforcement task (Vega Ruiz 2009).

*Intake procedures.* By “intake procedures,” I mean standard responses to inquiries and complaints. Are they welcomed or discouraged? Are they recorded consistently or haphazardly? Are they prioritized and, if so, how? The answers to these—and similar—

---

<sup>8</sup> Neither strategy guarantees probity, and both may backfire. See Bunker and Cohen (1983) on the potential affinity between teamwork and corruption among decentralized government personnel in Brazil, Olson (1993) on the dangers of “roving bandits,” and Spiller and Tommasi (2003) on the costs of personnel rotation more generally.

questions influence not only who is investigated but when and how and nonetheless vary markedly within and across agencies (see, e.g., GAO 2009).

*Target selection.* By “target selection,” I mean the decision criteria and procedures that determine which firms and enterprises are inspected in the first place. The literature tends to draw a first-order distinction between post hoc (i.e., complaint- or incident-based) and programmed (or proactive) inspections (Richthofen 2002; Weil 2008) and a second-order distinction between “random” assignment and “risk-based” targeting of the latter. “In this context,” explain Florentin Blanc and Marielle Leseur of the World Bank, “risk is understood as a combination of the probability and the scope of the hazard a business activity may create” (Blanc and Leseur 2011, p. 53). Latin American labor ministries pursue both post hoc and programmed inspections, as well as both types of assignment process, albeit in different degrees and combinations in different times and places (Jatobá 2002, p. 15; Vega 2009, pp. 20-3).

*Field methods.* By “field methods,” I mean the standard operating procedures and tactics the inspectors use in the field. Do they rely on interviews, visual inspections, or reviews of administrative records? Do they investigate a broad range of activity and behavior in an effort to uncover multiple violations or do they limit their investigations to discrete activities and behaviors? Do they offer advice and counsel or simply take stock of the situation at hand? Do they strive for “consistency of treatment” (von Richthofen 2002, p. 118) or tailor their efforts to the exigencies of the particular situation? The answers to these—and similar—questions are in part determined by law and statute (e.g., the breadth of the inspector’s jurisdiction) but are in part the product of administrative practices and customs.

*Remediation.* By “remediation,” I mean the response to perceived noncompliance in each particular case. The options available tend to include conciliation, formal as well as informal warnings (often with the stipulation of a rectification period), the provision of advice and counsel, and/or the deployment or initiation of administrative or judicial sanctions; on rare occasions, inspectors are allowed to issue ‘cease and desist’ orders or to embargo materials and equipment.

Bear in mind, however, that the personnel policies, intake procedures, targeting criteria, field methods, and remediation practices that are the building blocks of national (or organizational) enforcement strategies are, like Legos, subject to a virtually endless array of permutations. Which practices tend to be observed in contemporary Latin America? Are there affinities between different practices? And which permutations are worthy of further investigation? The answers will necessarily influence—and be influenced by—the available candidates for experimental treatment.

#### **4. Candidate experimental subjects**

Section 3 discussed five broad aspects of enforcement strategy that are potentially subject to experimental manipulation: personnel policies; intake procedures; targeting criteria; field methods; and remediation practices. On whom or what are the treatments likely to be imposed and experiments likely to be performed? The answer is by no means obvious, but five candidates immediately spring to mind: firms, jurisdictions, violators, inspectors, and complainants or complaints.

*Firms.* The most obvious candidates for treatment are individual firms, for the random assignment of inspections or inspection strategies to firms is a relatively straightforward process. But efforts to randomize by firm have at least three disadvantages. First, they are unrealistic. Real world inspectorates have to deal with complaints (and incidents) and are therefore unable to assign inspectors to firms at random. Second, they are considered inefficient. “Random checks are relatively easy and inexpensive to implement,” explain Blanc and Leseur, “but risk-based targeting remains a better option” (Blanc and Leseur 2011, p. 53). And, finally, they have trouble identifying the “demonstration effect” on nearby or networked firms. Inspections are designed not only to identify and correct noncompliance among the firms being investigated but to deter noncompliance among neighboring or known firms that learn about and fear the threat of inspection. A simple comparison of inspected firms and controls that have not been inspected is unable to capture the demonstration effect and might well suffer from contamination that would *underestimate* the effects of treatment (i.e., if the controls knew of the treatment and responded accordingly).

*Jurisdictions.* An alternative approach would randomly assign enforcement strategies (e.g., personnel policies, intake procedures, target selection criteria, field methods, etc.) to different jurisdictions (i.e., clusters)—rather than inspections or inspectors to different firms—in an effort to capture not only the direct effects of inspection on targeted firms but the broader effects of different strategies across the labor market as a whole. While cluster-based designs have the potential to capture the demonstration effect, they are likely to suffer from limited degrees of freedom and to be vulnerable to contamination in the Latin American context—where both the number of individual jurisdictions and the population of employers tend to be rather small and unevenly sized.<sup>9</sup>

*Violators.* A third approach would focus not on the labor market as a whole—by treating firms or jurisdictions—but on suspected violators by randomly assigning different responses to their transgressions. For instance, some might be given a warning, others might be sanctioned, and others might be given compliance-assistance measures in addition to a warning or in lieu of a sanction—with promises of greater sanctions in the event of future transgressions (see Ayres and Braithwaite 1992). While efforts to address similar violations in different ways would in all likelihood lend insight into the most effective means of preventing recidivism among known violators, they would yield little or no insight into deterrence or compliance writ large.

*Inspectors.* A fourth approach would randomly assign the same task (e.g., investigating complaints from workers) to inspectors of different types in an effort to assess the impact of personnel policies that are not susceptible to randomization. Consider, for example, the impact of merit-based recruitment and retention policies. While they are widely portrayed as indispensable to the performance of the public sector in the developing world, they are ill-disposed to experimentation for at least two reasons: first, they are generally subject to the control of the legislature or an independent civil service commission, and are thus

---

<sup>9</sup> See Cohen and Dupas (2010) and George et al. (2012) for a cluster-based design with comparably small numbers of clusters and Gerber and Green 2012 (pp. 80-84) on the advantages and disadvantages of clustering more generally.

invulnerable to random assignment by labor administrators; and second, they are believed to influence public sector performance by altering the “long-term career rewards” (Rauch and Evans 2000, p. 52) available to public servants, and are thus unlikely to be captured in a meaningful way by short-term (and time-bound) experimental interventions. But Latin American labor ministries frequently employ inspectors of different educational or professional backgrounds and/or under different contractual arrangements simultaneously (see, e.g., Schrank and Piore 2007; Bensusán 2008; Amengual 2011), and analysts might therefore gain insight into the effects of different personnel policies by assigning the same task to different *types* of inspector in the same country.

*Complaints and Complainants.* The final approach would randomly assign complaints or complainants of different types to inspectors in an effort to understand intake procedures and follow up. It would thus constitute a variant of the classic audit study (e.g., GAO 2009) and offer insight into the complaint process from the bottom up. Are inspectors more responsive to different types of complaints or complainants? If so how? And which ones? To what effect?

Of course, the five approaches are neither exhaustive nor mutually exclusive, and a packaged or modular approach may prove desirable. But they nonetheless offer a starting point from which to consider experimental assessments of enforcement strategies and their implications.

#### **4. Practical considerations**

Potential donors, analysts, practitioners, and partners should ask themselves at least three questions before joining forces to conduct a field experiment. Is the experiment likely to be useful? Is the experiment ethical? And is the experiment viable? The answers to these questions will depend upon a number of considerations and will determine the appeal of the proposed intervention. I will address each consideration in turn before introducing a checklist designed to facilitate the assessment of the aforementioned designs.

*Utility.* The first question concerns the *utility* of the proposed experimental intervention. Utility is gauged in terms of both the *credibility* and the *applicability* of the experiment’s lessons. We would therefore need to assess not only whether the experiment allowed for the credible estimate of causal (or treatment) effect (Gaines and Kuklinski 2011; Gerber and Green 2012) but also whether, if successful, the intervention lent itself to widespread adoption. An experiment that is decoupled from organizational needs and capacities is of questionable utility (Stoker 2010). Experiments on labor inspection systems are likely to confront two additional challenges that merit further discussion: the absence of valid data on compliance outcomes including violations; and the presence (and theoretical salience) of demonstration and spillover effects like the ones alluded to in the previous section. In the biomedical sciences that provide the implicit model for most experimental research, after all, the process of outcome measurement is relatively straightforward. Investigators have valid outcome measures at their disposal (e.g., tumor size, blood flow, survival rates, etc.) and subjects are at least theoretically willing to give the investigators any and all information that might contribute to their recovery. In the context of labor inspection, however, outcome measurement is rendered difficult by the proscribed or illicit nature of the outcomes themselves and the fact that subjects who are not in compliance with the law have a vested interest in avoiding detection. Independent surveys offer one possible

solution; administrative data provide another. For instance, Gustavo de Andrade and his colleagues carried out a survey *and* used administrative data in an effort to assess the effects of inspection on formalization in Brazil (Andrade et al. 2013), and their study also explores the possibility of spillover effects. While spillovers imply the transmission of treatment effects from subject to subject, and thus constitute a violation of the “stable unit treatment value assumption” (SUTVA) that underpins most randomized trials, they also imply deterrence and/or learning, and are thus policy relevant to the case of labor inspection. Andrade et al. explicitly address the possibility of spillovers by investigating the effects of inspections on neighboring firms, and Sinclair et al. have recently discussed designs for the assessment of SUTVA violations (Sinclair et al. 2012).

An evaluation of utility would demand answers to the following questions:

- Does the design address policy considerations highlighted by the logic model? The logic model not only highlights two different sources of regulatory noncompliance—ignorance and avarice—and two corresponding remedies—capacity-building and cost-benefit calculation—but implies that compliance outcomes will necessarily be influenced by organizational inputs and contextual factors regardless of diagnosis, and the designs should thus take the latter factors no less seriously than the former.
- Does the design allow for the valid estimation of a causal effect (i.e., counterfactual analysis)? The most demanding assessments of causal effects exploit not only the random assignment of treatments to subjects in an effort to isolate the impact of the former on the outcome of interest but the isolation and symmetrical treatment of the treatment and control groups in an effort to ensure that any observed differences in outcomes are due neither to the unintentional contamination of controls by treatments nor the uneven treatment of the two groups during the course of the experiment (Gerber and Green 2012, p. 41). “Random assignment is the analyst’s ‘protection’ against systematic differences in observable (and unobservable) characteristics that can introduce bias into the estimate of the true average causal effect,” explain David Konisky and Christopher Reenock. “In the absence of random assignment or large groups, systematic differences between the groups may emerge” (2013, p. 365).
- Does the design allow for the identification and differentiation of long-term as well as short-term effects? In his path-breaking work on labor inspection in Brazil, Roberto Pires draws a distinction between mere “compliance” with labor law and “sustainable compliance.” While the former tends to compromise competitiveness, and is therefore a source of attrition (Andrade et al. 2013) as well as backsliding into non-compliance, the latter tends to “reconcile compliance with economic efficiency” (Pires 2008, p. 222), and is therefore more lasting. Others have drawn similar distinctions between short-run and long-run compliance (see, e.g., Shapiro and Rabinowitz 1997), albeit not necessarily for the same reasons, and the ideal design would thus identify not only the immediate but the more enduring effects of inspection.
- Does the design allow for the estimate of spillover effects? The spillover effects that are a threat to validity in the classical experiment are an outcome to be investigated in the case of labor inspection, and the assumption of “non-interference” (i.e., SUTVA) is thus likely to be abandoned for designs that “randomly assign varying degrees of

secondhand exposure” (Gerber and Green 2012, p.282) in an effort to capture spillovers.

- Are valid compliance data available or potentially available at reasonable cost? The illicit nature of non-compliance with labor law poses a grave problem for experimental analysis, for the ultimate outcomes are difficult to observe and analysts will thus have to choose among costly surveys that are prone to systematic as well as random error, administrative data that are subject to biases of their own, and the investigation of policy relevant outcomes that are more easily measured than compliance (e.g., number of alleged violations identified, client satisfaction, etc.).
- Are the statistical underpinnings of the design well understood? The statistical demands of experimentation are generally simpler and better known than the demands of nonexperimental research, for the omitted variable problem that tends to bedevil the latter is addressed by means of design (Rubin 2008) or random assignment (Gerber and Green 2012, p. 95) in the former. But experimental results are rarely self-evident, and statistical analysis is not only necessary but in certain situations—e.g., scaling, noncompliance, weighting, etc.—quite complicated.
- Does the intervention lend itself to adoption if successful? Many potentially useful treatments are simply verboten for political or economic reasons, e.g., they demand too many resources, run contrary to the interests of powerful actors, or pose ethical concerns. While their impacts are potentially interesting, they are not worth pursuing in lights of their impracticality. ILAB’s resources would be better spent on more likely candidates.

*Ethical probity.* The second question concerns the ethical probity of the proposed experimental intervention. The ethics of a field experiment are assessed by considering the intervention’s effects not only upon the experimental subjects but upon the broader host community (see, e.g., Barrett and Carter 2010, p. 520), and the latter is a particular risk in the case of workplace regulation (Parker and Nielsen 2009, p. 58). Consider, for example, an audit study designed to evaluate inspector behavior by assigning testers with different characteristics (e.g., age, gender, employer, occupation, grievance, etc.) to lodge fictitious complaints against hypothetical employers (Pager 2003). While the results would tell us a good deal about how inspectors respond to different types of complaints and informants, and would therefore be highly useful, they would also distract inspectors from addressing real complaints involving genuine threats to actual workers, and would thus be ruled unethical by most oversight authorities. A design that neither increased the net burden on existing inspectors nor threatened the net resources available to aggrieved workers would be more likely to pass ethical muster.

- Does the design allow for the adequate protection of experimental or human subjects? The protection of human subjects is a non-negotiable aspect of experimental research and typically involves not only confidentiality but informed consent; that is, the disclosure of risks that are more severe than “the harms and discomforts ordinarily encountered in daily life” (Tilden 2008) to potential subjects and the solicitation of their active consent to participate on a voluntary basis. While Federal guidelines allow for the waiver of informed consent in the event that the research could not be carried

out otherwise, they ask the institutional review boards that govern research to consider the “susceptibility, vulnerability, resilience and experience” of the subject population when deciding whether to grant the waivers (Tilden 2008), and aggrieved workers in developing countries are likely to score higher on susceptibility and vulnerability than on resilience and experience.

- Does the design protect non-participants from harm? The principle of “do no harm” to research subjects is relatively well established in theory, if at times ambiguous in practice, but the principle of “do no harm” to non-participants is less familiar, if no less important, in the social scientific literature (Morton and Williams 2010, p. 473; Peyton 2012, p. 30; Barrett and Carter 2014, p. 61). For instance, Alan Gerber and Donald Green’s recently published textbook on field experiments includes an entire appendix on the protection of human subjects and nary a word on the protection of non-subjects who at times pay an indirect price for field experiments (Gerber and Green 2012).<sup>10</sup>

*Viability.* The third question concerns the *viability* of the proposed intervention. Are the key stakeholders willing and able to implement the relevant treatments? The answers will in all likelihood depend not only upon their cost, difficulty, and legality but upon the probity and competence of the inspectors themselves. If the inspectors are incompetent or unprincipled, and/or consider the treatments burdensome, they may shirk or refuse to put them into effect, thereby compromising the viability of the experiment. While statistical corrections are available, they offer an imperfect solution “when there is noncompliance among the treatment group” (Fuente and Whittington 2013, p 422; see also Barrett and Carter 2010, p. 523) and increase demands on the sample size at a minimum (Duflo et al. 2008).

- Is the proposed interventional legal? Many potentially interesting treatments are illegal (e.g., the intimidation of an employer, the subornation of an inspector, etc.). Like designs that are ill-disposed to adoption, they are not worthy of consideration.
- Is the proposed intervention financially feasible? Cost-effectiveness is another *de minimus* condition of a feasible intervention. Finite administrative costs would seem to be tolerable; ongoing capital and/or current expenditures would seem unacceptable.
- Are the relevant stakeholders willing and able to carry out the proposed intervention? The final criterion concerns the labor ministry’s *willingness* and *ability* to carry out the proposed intervention according to plan. While the former is likely to depend on the

---

<sup>10</sup> Risks to non-subjects would seem particularly salient in experiments on law enforcement given the nature of opportunity costs. Consider, for example, Fried et al.’s recent effort to see whether and under what conditions drivers pulled over for making illegal turns would be asked for bribes by Mexican police officers. While they assert that the illegal turns they induced were safe (Fried et al. 2010, p. 81), and ask the reader to take their assertion at face value, they simply ignore the costs to the non-participants in the study—whom the police officers may have been protecting had they not been chasing experimental traffic violators. See also Barrett and Carter’s treatment of Bertrand et al.’s effort to induce non-drivers in India “to bribe officials in order to receive a license without having successfully completed the required training and an obligatory driver safety examination” (Barrett and Carter 2010, pp. 519-20; see also Bertrand et al. 2007).

design’s cost (including opportunity cost) and utility, the latter is likely to depend on the ministry’s capacity and ILAB’s support, and in the absence of both willingness and ability on the ministry side, noncompliance is likely to prove overwhelming.<sup>11</sup>

The following checklist thus includes a series of questions by which to evaluate possible designs in terms of each of the three broad criteria. The questions can be answered “yes,” “no,” or “not applicable” (NA) for each of the designs, and the results can be used to determine whether the design is recommended or not.

**Table 1: Feasibility checklist for experimental designs**

	Question	Answer			Recommended	
		Yes	No	NA	Yes	No
<b>Utility</b>	Does the design address policy considerations highlighted by the logic model?					
	Does the design allow for the valid estimation of a causal effect (i.e., counterfactual analysis)?					
	Does the design allow for the identification and differentiation of long-term and short term effects?					
	Does the design allow for the estimate of spillover effects?					
	Are valid compliance data available or potentially available at reasonable cost?					
	Are the statistical underpinnings of the design well understood?					
	Does the intervention lend itself to adoption if successful?					
<b>Morality</b>	Does the design allow for the adequate protection of experimental or human subjects?					
	Does the design protect non-participants from harm?					
<b>Viability</b>	Is the proposed interventional legal?					
	Is the proposed intervention financially feasible?					
	Are the relevant stakeholders willing and able to carry out the proposed intervention?					

### 5. Candidate experimental designs.

Which designs merit consideration in light of the aforementioned concerns? I will discuss three candidates: a comparison of responses to complaints by professional inspectors and partisan recruits; a comparison of risk-based targeting and the randomized assignment of programmed inspections by their supervisors; and a comparison of sanctions and support to known violators in an effort to bring about sustainable compliance. The three designs

<sup>11</sup> Intent-to-treat type models and local average treatment effect (LATE) estimators are available in the event of noncompliance, but they are likely to be overwhelmed by widespread opposition, inefficacy, or corruption among ministry officials. In fact, Barrett and Carter note that the “implementer compliance” problem not only “creeps in earlier in the research” process but compromises the integrity of the entire data collection effort (p. 523). “Indeed, when research is subcontracted to implementing partners,” they argue, “study authors commonly do not even know if such sampling bias exists in the data.”

address three of the most important questions confronting contemporary labor ministries: Who should conduct their inspections? Whom should they inspect? And how should they respond to violations in the event that they are detected?

*Professionals v. Partisans.* Figure 4 drew a distinction between the minority of countries that appoint their inspectors on the basis of “strict selection criteria” (Vega 2009, p. 25) and the more common Latin American practice of “political appointments” (p. 25). While recruitment practices are among the “most important policies” (Richthofen 2002, p. 109) established by the labor ministry, and necessarily shape everything that follows, they are immune from experimental manipulation for legal as well as practical reasons alluded to in Section 2, i.e., they are rarely determined by the labor ministry alone, and they are unlikely to have their predicted effect if they are manipulated on a temporary basis. After all, the principal link between professionalization and probity is an *esprit de corps* that produces “a sense of commitment to corporate goals” (Rauch and Evans 2000, p. 52) at the base of the organization, and the latter is unlikely to be encouraged by a *temporary* process of professionalization.

But the simultaneous presence of professionals and partisans in the same agency affords potential experimenters a different source of insight into the effects of personnel policies. By randomly assigning worksites (in the event of programmed inspection) or complaints (in the case of post hoc investigations) to different types of inspectors, after all, they can potentially trace any differences in outcomes to the differences among the inspectors and derive corresponding conclusions for improved personnel policies.

The baseline specification is distilled into equation 1.

- (1)  $y_{ij} = a + b_T T_j + e_{ij}$ , where  $y$  is the outcome of interest (e.g., compliance),  $a$  is a constant,  $T$  is a dummy variable that tracks the presence or absence of the treatment (e.g., professional = 1; partisan = 0) for inspector  $j$ ,  $e$  is an error term, and subscript  $i$  indexes the individual inspection.

The more exacting specification in equation 2 includes a vector of control variables ( $X$ ) for individual inspectors—who are likely to differ in a number of ways above and beyond their levels of professionalization (e.g., education, experience, gender)—and a vector of control variables ( $Z$ ) for different types of worksites ( $k$ ) that are likely to differ in terms of size, sector, ownership, and the like.

- (2)  $y_{ij} = a + b_T T_j + bX_{ij} + bZ_{ik} + e_{ijk}$

Consider, for example, complaints. While they are frequently assigned to inspectors of different types in a more or less arbitrary manner (Jatobá 2002), they could easily be assigned to inspectors of different types in a formally random manner (e.g., a treatment group of inspections assigned to professionals and a control group of inspections assigned to partisans), and any observed differences in their disposition—i.e., response times, remediation strategies, compliance outcomes—could thus be attributed to differences in type of inspector. The differences in treatment would demand careful tracking, of course, but the obstacles to their assessment are by no means insurmountable. For instance, response time can be tracked by the ministry itself with administrative data. How many

complaints do inspectors of different types address per month? Formalization can be tracked with administrative data as well. Are firms that are visited by certain types of inspectors more likely to be registered in the following year? Inspection practices can be tracked by means of inspection reports of debriefings. What exactly did the inspectors of different types do on each visit? And more refined data can be obtained by means of follow-up surveys of firms and/or workers. Are firms that were visited by certain types of inspectors more likely to remain compliant in the aftermath of the visit? Are workers whose complaints were addressed by certain types of inspectors more likely to express satisfaction with the process after the fact?

The “professionals v. partisans” design offers answers to these questions not only for the registered enterprises that are already relatively well understood in Latin America but for the informal enterprises that tend to fly under the radar. After all, the design randomly assigns complaints to inspectors regardless of whether they come from informal or formal sector workers, and thereby allows the ministry to both (i) treat informal as well as formal sector employers with different types of inspectors and (ii) use administrative data on registration as more or less objective indicators of their responses to treatment. If firms that are assigned to professionals are more likely to register (or stay registered) than firms that are assigned to their politically appointed counterparts, for instance, the professionals have presumably achieved better compliance outcomes than the partisans. Of course, the ministry may want to collect additional data in an effort to understand specific *patterns* of compliance or stakeholder *perceptions* of the process, but the mere combination of administrative data on registered firms and complaints about the behavior of their unregistered counterparts makes the process of outcomes assessment that much easier.

The real question therefore concerns the availability of policy relevant differences among inspectors working in the same agency. Are the inspectors sufficiently heterogeneous to allow for experimentation? The answer would seem to be affirmative. For instance, the Dominican Republic employs not only lawyers who have been incorporated into the civil service by means of competitive examinations and partisan recruits who are devoid of civil service protection but partisan recruits who lack law degrees but have nonetheless passed the exam and have thus been incorporated into the civil service (Schrank 2009). The island nation thus offers variation on credentials as well as civil service protections and makes an ideal spot for experimentation on personnel policies.

Other countries offer variation along different professional axes. For instance, Mexico employs not only lawyers but engineers recruited by means of competitive exams as well as partisan incumbents with little or no education to speak of, and thus offers a broader range of variation on credentials as well as civil service protections (Bensusán 2008, pp. 107-8). Costa Rica stipulates that inspectors have degrees “in law, social science, labour administration or a related field” (Vega 2009, p. 25), and thus allows for the comparison of different types of academic credentials in the same social and economic environment. Guatemala employs inspectors with legal training (if not always law degrees) as well as partisan recruits, but incorporates neither into the civil service (Schrank and Piore 2007), and thus offers insight into the effects of credentials without protections. Ecuador has recruited more than one hundred lawyers in recent years in a deliberate effort to create a

meritocratic labor inspectorate (MRL 2011, p. 6).<sup>12</sup> And El Salvador and Honduras are in the process of consolidating their civil services (Vega 2012, p. 7), and thus offer distinct possibilities for experimentation in real time.

*Risk-Based Targeting v. Random Assignment.* The first design assigns worksites and/or complaints to inspectors of different types. It is illustrated by way of reference to post hoc investigations, and thus looks at ex post facto outcomes (i.e., in response to reported problems). While post hoc investigations constitute an indispensable response to tragedy or abuse, and are therefore an inescapable part of the inspector's job, they are a less efficient means of control than programmed inspections (Richthofen 2002, p. 90; Fenwick et al. 2007, p. 41; Weil 2008, p. 350), for complaints "tend to occur long after the fact and do not necessarily indicate the true state of the workplace in any event—especially in developing countries, where ignorance and illiteracy are, in all likelihood, responsible for 'false positives' as well as 'false negatives'" (Schrank 2009, p. 99). Child labor constitutes a clear example. Neither children nor their parents tend to "report instances of child labor," explains Wolfgang von Richthofen of the ILO, "because their circumstances are such that they see no alternative and therefore have an interest in perpetuating it. They may not even know whether there are laws against it" (Richthofen 2002, pp. 221-22).

The best approach to programming is by no means obvious, however, for experts tend to dispute the relative merits of randomized and risk-based alternatives. While random inspections "are relatively easy and inexpensive to implement" (Blanc and Leseur 2011, p. 53), they arguably squander local knowledge that might otherwise be put to good use in selecting high risk targets.

A second design would therefore randomly assign inspectors to programmed inspections by means of risk-based targeting, on the one hand, and the random selection of targets, on the other, in an effort to understand the relative costs and benefits of the two approaches.

The baseline model is distilled into equation 3.

- (3)  $y_i = a + b_T T + e_i$ , where  $y$  is the outcome of interest (e.g., noncompliance),  $a$  is a constant,  $T$  is a dummy variable that tracks the presence or absence of the treatment (e.g., risk-based = 1; random = 0),  $e$  is an error term, and subscript  $i$  indexes the individual observation.

Random assignment should ensure the comparability of the treatment and control groups net of treatment. But control variables can be added to improve precision and account for stochastic differences across inspectors or worksites.

$$(4) y_{ij} = a + b_T T_j + bX_{ij} + bZ_{ik} + e_{ijk}$$

Are risk-based inspections really more likely to reveal noncompliance than random ones? Do the added costs of risk assessment outweigh the benefits? Does the targeting strategy

---

<sup>12</sup> Additional information on the Ecuadorean reforms was gleaned from anonymous interviews carried out in Quito in March 2014.

influence the type—or simply the volume—of noncompliance unearthed by the inspectors? Is either approach more conducive to remediation? And might deterrence vary by target selection process as well? The answers are likely to depend not only on the definition and measurement of risk—i.e., the probability of noncompliance, magnitude of harm, or a combination of the two (Blanc and Leseur 2011, p. 8)—but on the approach to random assignment. Will ministry officials draw probability samples from existing lists of registered firms, use nonprobability methods designed to sample “hard to reach” populations (e.g., capture-recapture, respondent-driven, and time-space sampling) in an effort to capture informal as well as formal enterprises (Marpsat and Razafindratsima 2010), allow inspectors to choose their own targets, or use multiple random assignment schemes in an effort to provide a “fair test” (Gerber and Green 2012, p. 7) of the various possibilities? The answers are decidedly important, for they determine not only the cost but the composition of the control group against which risk-based targeting is judged.<sup>13</sup>

Outcome measurement is likely to prove difficult, however, no matter which sampling scheme is used. While inspectors deployed to address specific risks and their randomly assigned colleagues are likely to encounter different rates and patterns of noncompliance *ex ante*, and to address perceived violations in different ways, their supervisors will lack objective data on compliance outcomes—beyond rates of formalization that might be artifacts of the sampling schemes by which their targets are selected in the first place—and will thus be left to wonder whether the apparent differences are real or imagined in the absence of follow-up surveys that are likely to prove costly. Of course, the need for a follow-up survey need not prove disqualifying; on the contrary, it is an all but inherent concomitant of research on illicit activities that do not speak for themselves. But the very illegality that renders follow-up data collection necessary also renders it more costly and difficult, and constitutes a real threat to the assessment of the experimental treatment.

*Carrots v. Sticks.* The logic model suggests that noncompliance is likely to result from ignorance as well as avarice, and that the treatment should reflect the diagnosis. Where noncompliance is willful, for example, sanctions are in all likelihood the best way to alter the employer’s cost-benefit calculus in an effort to bring about compliance. Where the problem results from ignorance, however, sanctions are less promising (or perhaps even counterproductive) and carrots—like compliance assistance—might therefore prove more effective than sticks.

What would compliance assistance look like? The existing literature discusses two broad types of support—legal and administrative versus technical and productive—and two main delivery channels: informal and formal (see Figure 5).

**Figure 5: Potential patterns of compliance assistance**

		Structure	
		<i>Informal</i>	<i>Formal</i>
Content	<i>Legal and administrative</i>	Informal legal advice	Formal legal advice
	<i>Technical and productive</i>	Informal technical or productive advice	Formal technical or productive support

<sup>13</sup> Anecdotal evidence suggests that risk-based targeting can be quite effective. See, e.g., El Nuevo Diario (2009).

The most common forms of compliance assistance are imparted informally by inspectors. For instance, Matthew Amengual recently found that most of the inspectors he surveyed in Argentina offered managers advice about their legal obligations as a matter of course, and that many more made suggestions about their productive practices as well (Amengual 2014). Nor are his findings exceptional. On the contrary, a substantial body of literature implies that the knowledge and proclivity to impart advice are themselves hardwired into inspectors—and general-purpose inspectors in particular—who “are exposed to a wide range of business practices” in the course of carrying out their daily responsibilities and are “thus able to play the role of business consultant by transferring best practices from leading to lagging enterprises” (Piore and Schrank 2008, p. 7; see also Pires 2008). But the most sophisticated Latin American labor ministries have formalized the process of pedagogy in a number of innovative ways. For instance, Chile allows noncompliant firms that employ fewer than 50 workers to undergo training in lieu of paying fines (Piore and Schrank 2008, p. 7). And the Dominicans have not only linked inspection to training in a number of different ways but have given their inspectors a tacit role in subsidy allocation as well (Schrank 2013, p. 9).

Are carrots likely to contribute to compliance? The answer is by no means obvious (see, e.g., Rosado Marzán 2012), and an experimental evaluation would therefore seem to be in order. Andrade et al.’s recent experiment on formalization practices in Brazil offers a template. They randomly assigned informal enterprises to “a control group or one of four treatments groups: the first received information about how to formalize; the second received this information and free registration costs along with the use of an accountant for a year; the third group was assigned to receive an enforcement visit from a municipal inspector; while the fourth group was assigned to have a neighboring firm receive an enforcement visit to see if enforcement has spillovers” (Andrade et al. 2013, p. 1). While inspection visits triggered significant increases in formalization, information, subsidies, and spillovers had either insignificant or—in some cases—negative effects.

But Andrade et al. look only at formalization, and not at noncompliance more generally, and fail to examine the potential impact of technical and productive assistance that might allow noncompliant firms to reconcile their competitive needs with their contractual and legal obligations. A third design might therefore randomly assign noncompliant firms more generally to a control group and one of three treatment groups: the first would be threatened with sanctions; the second, would be introduced to public training authorities and/or development banks; and the third would be threatened with sanctions, introduced to public training authorities and/or development banks, and offered a rectification period in the event that they agreed to work with the latter in an effort to find positive sum solutions for their workers and their bottom lines. A follow-up survey would be carried out over the course of the next year or more—budget permitting—to assess whether and to what degree the various treatments fostered sustainable compliance.

The baseline model is distilled into equation 5.

$$(5) y_i = a + b_T T + e_i,$$

where  $y$  is the outcome of interest (e.g., compliance),  $a$  is a constant,  $T$  is a dummy variable that tracks the presence or absence of the treatments (e.g., threatened sanctions, introduction to public training authorities

and/or banks, and a combination of the two),  $e$  is an error term, and subscript  $i$  indexes the individual observation.

Given the likelihood of noncompliance with the more complicated protocol in the third design, an intention-to-treat or local average treatment effect approach might be in order (Duflo et al. 2007). But a follow-up survey would almost certainly be required to evaluate the effects of treatment regardless of the chosen estimator. After all, Andrade et al. used both administrative data and a follow up survey; however, they were interested in formality alone and not compliance more generally—let alone whether compliance had been facilitated by productive upgrading.

Table 2 summarizes the key features of each proposed design including the treatment to be imposed, hypothesized outcomes and impacts, approach to counterfactual estimation, data requirements, disaggregated timeline, human subject and non-subject protections, regulatory requirements, candidate countries, and additional considerations.

**Table 2: Summary of proposed designs**

	<b><i>Design 1: Professionals v. partisans</i></b>	<b><i>Design 2: Risk-based v. random targeting</i></b>	<b><i>Design 3: Carrots v. sticks</i></b>
<i>Treatment</i>	Visits by professional inspectors	Risk-based targeting of programmed inspections	Technical or productive aid (and perhaps threats)
<i>Expected outcome</i>	More efficient and honest inspections	More efficient inspections	Productive upgrading into compliance
<i>Potential short-run impact</i>	Formalization and improved compliance	More violations revealed per inspections visit	Improved compliance
<i>Potential long-run impact</i>	Spillovers to neighboring firms	Remediation and compliance	Sustainable compliance
<i>Counterfactual</i>	Inspection by incumbent or partisan inspectors	Random assignment of programmed inspections	Threats alone
<i>Data needs and methods</i>	Administrative data and/or follow-up survey	Administrative data and a follow up survey	Administrative data and a follow up survey
<i>Design timeline</i>	1-2 months	3-6 months	3-6 months
<i>Implementation timeline</i>	6-12 months	6-12 months	6-12 months
<i>Data collection timeline</i>	6-12 months	6-12 months	12-24 months
<i>Data analysis timeline</i>	1-2 months	1-2 months	1-2 months
<i>Protection of subjects</i>	Confidentiality	Confidentiality	Confidentiality
<i>Protection of non-subjects</i>	Resource diversion (low)	Resource diversion (medium)	Resource diversion (high)
<i>Regulatory requirements</i>	Heterogeneous labor inspectors	Relatively capable inspectors	Skilled inspectors and industrial policy agencies
<i>Candidate countries</i>	DR, Mexico, Ecuador, most of Central America	Brazil, Chile, CR, DR, Uruguay, perhaps Argentina	Brazil, Chile, CR, DR, Uruguay, perhaps Argentina
<i>ILAB cost</i>	Approximately \$250,000	Approximately \$300,000	Approximately \$500,000
<i>Country cost</i>	Variable	Approximately \$50,000/yr	Variable
<i>Additional concerns</i>	A successful experiment would have potentially costly policy implications (i.e., professionalizing the labor inspectorate)	Outcome measurement is difficult; nonresults are consistent with poor targeting as well as the null hypothesis of no effect	Outcome measurement is difficult; nonresults are consistent with inadequate technical/productive support as well as the null; support is costs; risk of resource diversion to least deserving firms

While all three designs have been formulated with economic and political constraints in mind, they are by no means easy to implement, evaluate, or justify on ethical grounds. Implementation will prove difficult in the absence of host country buy-in, for example, and evaluation will prove difficult in light of the illicit nature of the principal outcome of interest. Noncompliant firms have a vested interest in escaping detection, after all, and we lack a valid source of data on their behavior. While follow-up surveys offer one potential source, they are neither inexpensive nor foolproof, and at their best will constitute the weak observational link in the otherwise robust randomized approach. Similarly, the risk to experimental subjects would appear to be low, at least insofar as confidentiality can be preserved. But the risk to non-participants is potentially high, especially if resources are diverted from tasks that are valuable—if hard to estimate with the precision demanded by purists—to experimental interventions that may ultimately prove worthless.

The point is not to dismiss the experimental approach outright, however, but to instead recognize that, like all other approaches, it has costs as well as benefits. We are long past the days when experiments could be portrayed as silver bullets and must instead embrace the messy world in which they are but one of many potentially useful tools with no claim to privileged status (Ravallion 2009; Barrett and Carter 2010; Deaton 2010). While experiments are designed to produce “small but sure improvements” (Ravallion 2012, p. 103) in development policymaking, they have trouble addressing structural constraints that are not only important in their own rights but at times pose threats to the experiments themselves (Barrett and Carter 2010; Pritchett 2013).

## **Conclusion**

This paper has endeavored to do five things: first, introduce the Latin model of labor inspection and distinguish it from approaches found in other parts of the world; second, present a logic model linking labor law enforcement strategies to compliance outcomes in the region; third, discuss the component parts of workplace inspection and their potential susceptibility to experimental manipulation; fourth, develop a feasibility checklist with which to assess potential experimental designs; and, fifth, discuss potential designs and their likely costs and benefits. While the designs are likely to lend insight into the returns to different inspection strategies, they are likely to confront at least three pitfalls that are all but intrinsic to field experiments on the regulation of proscribed activity by porous bureaucracies and that have nonetheless received inadequate attention in the literature to date: non-implementation by inspectors or their supervisors (see, e.g., Pritchett 2013); the fact that “the ‘hard evidence’ from the randomized evaluation has to be supplemented with lots of soft evidence before it becomes usable” (Rodrik 2008, p. 5); and the risks posed to non-subjects in a context of public sector resource scarcity in particular (Barrett and Carter 2010). None of these pitfalls necessarily militates against the implementation of field experiments. But they serve as a useful reminder that experimentation is difficult under the best of circumstances and particularly difficult in developing country contexts where it is likely to be most useful. Randomized trials, in particular, may constitute the ‘gold standard’ of contemporary research practice, but as Barrett and Carter note: the original gold standard failed (2010, p. 524). We should be alert to the possibility that its metaphorical and methodological successor will as well.

## References

- Amengual, Matthew. 2011. *Enforcement without Autonomy: The Politics of Labor and Environmental Regulation in Argentina*. PhD dissertation. Department of Political Science. MIT.
- Amengual, Matthew. 2014. "Pathways to Enforcement: Labor Inspectors Leveraging Linkages with Society in Argentina." *Industrial and Labor Relations Review*. 67 (1): 2014.
- Andrade, Gustavo Henrique de, et al. 2013. "A Helping Hand or the Long Arm of the Law? Experimental evidence on what governments can do to formalize firms." World Bank Policy Research Working Paper 6435. Washington: World Bank.
- Ayres, Ian, and John Braithwaite. 1992. *Responsive Regulation: Transcending the Deregulation Debate*. Oxford: Oxford University Press.
- Banerjee, Abhijit, et al. 2012. "Can Institutions be Reformed from Within? Evidence from a Randomized Experiment with the Rajasthan Police." MIT Poverty Lab.
- Barrett, Christopher, and Michael Carter. 2010. "The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections." *Applied Economic Perspectives and Policy* 32 (4): 515-48
- Barrett, Christopher, and Michael Carter. 2014. "A Retreat from Radical Skepticism: Rebalancing Theory, Observational Data, and Randomization in Development Economics." Pp. 58-77 in D. Teele, ed., *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*. New Haven: Yale University Press.
- Bensusán, Graciela. 2008. *Regulaciones Laborales, Calidad de los Empleos y Models de Inspección: México en el Contexto Latinoamericano*. Mexico City: CEPAL.
- Bensusán, Graciela. 2009. "La Inspección de Trabajo en América Latina: teorías, contextos, y evidencias." *Estudios Sociológicos*. XXVII (81): 989-1040.
- Bertrand, Marianne, et al. 2007. "Obtaining a Driver's License in India: An Experimental Approach to Studying Corruption." *Quarterly Journal of Economics*. 122(4): 1639-76.
- Bhorat, Haroon, and Benjamin Stanwix. 2013. "Minimum Wage Enforcement in the Developing World." DPRU Policy Brief. University of Cape Town. May.
- Bizberg, Ilán. 2012. "Types of Capitalism in Latin America." Unpublished manuscript. El Colegio de México.
- Blanc, Florentin, and Marielle Leseur. 2011. *How to Reform Business Inspections: Design, Implementation, Challenges*. Washington: World Bank.
- Bunker, Stephen, and Lawrence Cohen." 1983. "Collaboration and Corruption in Two Colonization Projects: Toward a General Theory of Official Corruption." *Human Organization*. 42 (Summer): 106-14.
- Casale, Giuseppe, et al. 2012. "The Effectiveness of Labour Law and the Role of Labour Inspection." Paper presented to the 20<sup>th</sup> World Congress of the International Society for Labour and Social Security Law. 25-28 September.
- Castillo, Edith. 2013. "De no cumplir, las multas llegan a 5 mil salarios mínimos." Milenio. 18 noviembre. [http://www.milenio.com/region/cumplir-multas-llegan-salarios-minimos\\_0\\_192580769.html](http://www.milenio.com/region/cumplir-multas-llegan-salarios-minimos_0_192580769.html)
- Cohen, Jessica, and Pascaline Dupas. 2010. "Free Distribution or Cost-Sharing? Evidence from a Randomized malaria Experiment." *Quarterly Journal of Economics*. CXXXV (1): 1-45.

- Davis, Kevin, and Mariana Mota Prado. 2014. "Law, Regulation, and Development." Pp. 204-21 in B. Currie-Alder et al., eds., *International Development: Ideas, Experience, and Prospects*. Oxford: Oxford University Press.
- Deaton, Angus. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature*. 48 (June): 424-55.
- Duflo, Esther, et al. 2008. "Using Randomization in Development Economics Research: A Toolkit." Ch. 61 in T. Paul Schultz and John Strauss, eds., *Handbook of Development Economics, Vol. 4*. Amsterdam: North Holland.
- Duflo, Esther, et al. "Truth-telling by Third Party Auditors and the Response of Polluting Firms: Experimental Evidence from India." *Quarterly Journal of Economics* 128 (4): 1499-1545.
- El Nuevo Diario. 2009. "Trabajo inspecciona 1,165 bancas, bares y restaurants." 19 febrero. <http://www.elnuevodiario.com.do/app/article.aspx?id=140651>
- European Federation of Public Service Unions. 2012. *A mapping report on Labour Inspection Services in 15 European Countries: A SYNDEX report for the European Federation of Public Service Unions (EPSU)*. Brussels: EU Secretariat.
- Fenwick, Colin, et al. 2007. *Labour and Labour-Related Laws in Micro and Small Enterprises: Innovative Regulatory Approaches*. ILO SEED Working Paper 81.
- Fienberg, Stephen, et al. 2003. "Expert statistical testimony and epidemiological evidence: the toxic effects of lead exposure on children." *Journal of Econometrics*. 113: 33-48.
- Fried, Brian, et al. 2010. "Corruption and Inequality at the Crossroad: A Multimethod Study of Bribery and Discrimination in Latin America." *Latin American Research Review*. 45 (1): 76-97.
- Fuente, David and Dale Whittington. 2013. "Randomized control trials: what are they, why are they promoted as the gold standard for causal identification, and what can they (not) tell us." Ch. 31 in E. Araral et al., eds., *Routledge Handbook of Public Policy*. New York: Routledge.
- Gaines, Brian, and James Kuklinski. 2011. "Treatment Effects." Ch. 31 in J. Druckman et al., eds. *Cambridge Handbook of Experimental Political Science*. Cambridge: Cambridge University Press.
- Gerber, Alan, and Donald Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W.W. Norton.
- George, Christine, et al. 2012. "A Cluster-Based Randomized Controlled Trial Promoting Community Participation in Arsenic Mitigation Efforts in Singair, Bangladesh." *Environmental Health*. 11: 41-62.
- Godínez Vargas, Alexander. 2011. *Estudio sobre infracciones y sanciones en el ámbito de la inspección del trabajo: el caso de Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua y República Dominicana*. Geneva: OIT.
- Government Accountability Office. 2009. *Wage and Hour Division Needs Improved Investigative Processes and Ability to Suspend Statute of Limitations to Better Protect Workers Against Wage Theft*. Washington: GAO.
- Guzman, Hugo. 2011. "Trabajadores: los datos de la precariedad." *ALAI-América Latina en Movimiento*. <http://alainet.org/active/46372&lang=es>
- ILAB (Bureau of International Labor Affairs). 2013. *Experimental Design Strategies for Improving Labor Inspection Systems*. Performance Work Statement. Washington: Department of Labor.

- International Labour Office. 2006. *Business environment, labour law and micro- and small enterprises*. Governing Body, Committee on Employment and Social Policy, 297th Session, 2006. Geneva.
- Jatobá, Vera. 2002. *Labour Inspection within a Modernized Labour Administration*. Lima: ILO.
- Kagan, Robert, and John Scholz. 1984. "The 'Criminology of the Corporation' and Regulatory Enforcement Strategies." Pp. 67-95 in Keith Hawkins and John Thomas, eds., *Enforcing Regulation*. Boston: Kluwer-Nijhoff.
- Konisky, David, and Christopher Reenock. 2013. "Case Selection in Public Management Research: Problems and Solutions." *Journal of Public Administration Research and Theory*. 23: 361-93.
- Latinobarometer. 2005. *Latinobarometro 2005*. Octubre.
- Marpsat, Maryse, and Nicolas Razafindratsima. 2010. "Survey methods for hard-to-reach populations: introduction to the special issue." *Methodological Innovations Online*. 5 (2): 3-16.
- Melis, Christian. 2006. "Fiscalización y la Dirección del Trabajo." *Economía y Negocios*. [http://www.economiaynegocios.cl/mis\\_finanzas/detalles/detalle\\_emp.asp?id=166](http://www.economiaynegocios.cl/mis_finanzas/detalles/detalle_emp.asp?id=166)
- Melo, Alberto, and Andrés Rodríguez-Clare. 2006. *Produce Development Policies and Supporting Institutions in Latin America and the Caribbean*. Washington: IDB.
- Ministerio de Relaciones Laborales. 2011. *Rendición de Cuentas 2010-2011*. Quito: MRL.
- Moehler, Devra. 2010. "Democracy, Governance, and Randomized Development Assistance." *Annals of the American Academy of Political and Social Science*. 628 (March): 30-46.
- Morton, Rebecca, and Kenneth Williams. 2010. *Experimental Political Science and the Study of Causality*. New York: Cambridge University Press.
- Olson, Mancur. 1993. "Dictatorship, Democracy, and Development." *American Political Science Review*. 87 (3): 567-76.
- Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108 (5): 937-75.
- Parker, Christine, and Vibeke Nielsen. 2009. "The Challenge of Empirical Research on Business Compliance in Regulatory Capitalism." *Annual Review of Law and Social Science*. 5: 45-70.
- Peyton, Kyle. 2012. "Ethics and Politics in Field Experiments." *The Experimental Political Scientist*. 3 (1): 20-37.
- Piore, Michael, and Andrew Schrank. 2008. "Toward managed flexibility: The revival of labor inspection in the Latin world." *International Labour Review*. 147: 1-23.
- Pires, Roberto. 2008. "Promoting sustainable compliance: Styles of labour inspection and compliance outcomes in Brazil." *International Labour Review*. 147 (2-3): 199-229.
- Pritchett, Lant. 2013. "Folk and the Formula: Fact and Fiction in Development." UNU-WIDER Annual Lecture.
- Rauch, James, and Peter Evans. 2000. "Bureaucratic structure and bureaucratic performance in less developed countries." *Journal of Public Economics*. 75: 49-71.
- Ravallion, Martin. 2009. "Should the Randomistas Rule?" *Economists' Voice*. February: 1-5.
- Ravallion, Martin. 2012. "Fighting Poverty One Experiment at a Time: A Review of Abhijit Banerjee and Esther Duflo's *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*." *Journal of Economic Literature*. 50 (1): 103-14.

- Rodrik, Dani. 2008. "The New Development Economics: We Shall Experiment, But How Shall We Learn?" Unpublished manuscript. Princeton University.
- Rosado Marzán, César, 2012. "Punishment and Work Law Compliance: Lessons from Chile." *Hofstra Labor & Employment Law Journal*. 29 (2): 343-405.
- Rubin, Donald. 2008. "For Objective Causal Inference, Design Trumps Analysis." *Annals of Applied Statistics*. 2 (3): 808-40.
- Schrank, Andrew. 2009. "Professionalization and Probity in a Patrimonial State: Labor Inspectors in the Dominican Republic." *Latin American Politics & Society*. 51 (2): 91-115.
- Schrank, Andrew. 2013. "From disguised protectionism to rewarding regulation: The impact of trade-related labor standards in the Dominican Republic." *Regulation & Governance*. 7 (3): 299-320.
- Schrank, Andrew, and Michael Piore. 2007. *Norms, regulations, and labor standards in Central America*. Mexico City: CEPAL.
- Shapiro, Sidney, and Randy Rabinowitz. 1997. "Punishment versus Cooperation in Regulatory Enforcement: A Case Study of OSHA." *Administrative Law Review*. 49 (4): 716-32.
- Sinclair, Betsy et al. 2012. "Detecting Spillover Effects: Design and Analysis of Multilevel Experiments." *American Journal of Political Science* 56 (4): 1055-69.
- Spiller, Pablo, and Mariano Tommasi. 2003. "The Institutional Foundations of Public Policy: A transactional Approach with Applications to Argentina." *Journal of Law, Economics, and Organization*. 19 (2): 281-306.
- Stoker, Gerry. 2010. "Translating Experiments into Policy." *Annals of the American Academy of Political and Social Science* 628 (March): 47-58.
- Tendler, Judith. 2001. "The Informal Sector, Small Firms, and the Devil's Deal." *UNRISD News*. 24 (Spring/Summer): 9-10.
- Tilden, Samuel. 2008. "Secretary's Advisory Committee on Human Research Protections Letter to HHS Secretary." <http://www.hhs.gov/ohrp/sachrp/sachrletter013108.html>
- Ugarte Cataldo, José Luis. 2008. "Inspección de trabajo en Chile: Vicisitudes y Desafíos." *Revista Latinoamericana de Derecho Social*. 6: 187-204.
- Vaughan, Diane. 1998. "Rational Choice, Situated Action, and the Social Control of Organizations." *Law and Society Review*. 32 (1): 23-61.
- Vega Ruiz, Maria Luz. 2009. *Labour administration: To ensure good governance through legal compliance in Latin America. The central role of labour inspection*. Geneva: ILO.
- Vega Ruiz, Maria Luz. 2012. *Los procesos de selección y formación de los inspectores de trabajo: prácticas, programas y lagunas*. Geneva: ILO.
- Vega Ruiz, Maria Luz, and René Robert. 2013. *Labour Inspection Sanctions: Law and practice of national labour inspection systems*. Geneva: ILO.
- Von Richthofen, Wolfgang. 2002. *Labour Inspection: A Guide to the Profession*. Geneva: ILO.
- Weil, David. 2008. "A Strategic Approach to Labor Inspection." *International Labour Review*. 147 (4): 349-75.
- World Bank. 2013. *World Development Indicators*. Online.