The Pearson Institute Discussion Paper No. 2024-12

Statebuilding in the City: An Experiment in Civilian Alternatives to Policing

Christopher Blattman
Gustavo Duncan
Benjamin Lessing
Santiago Tobon

THE PEARSON INSTITUTE
FOR THE STUDY AND RESOLUTION OF GLOBAL CONFLICTS

Statebuilding in the City:
An Experiment in Civilian Alternatives to Policing

July 31, 2024

Abstract

State penetration varies widely within cities, with well-governed areas abutting persistently neglected ones. Governments are seeking ways to improve penetration, local security, and state legitimacy. We experimentally evaluate a 20-month non-police intervention in Medellín, Colombia, that dramatically increased municipal personnel and agency attention to 40 neighborhoods. Despite the intensity, average impacts on security and perceived legitimacy were negligible. Prespecified subgroup analysis reveals important heterogeneity, however. Where state governance began relatively lower, impacts were null to negative, but in initially high-governance sectors, security and state legitimacy significantly improved. These divergent impacts apparently resulted from city staff and agencies systematically underdelivering in low-initial-governance sectors. Bureaucratic capacity and incentives to deliver often depend on baseline state engagement, trust, and accountability. This could result in increasing marginal returns to statebuilding, which in turn would lead to persistent "neglect traps"—political attention and investment where state presence is already robust, reinforcing existing disparities.

1 Introduction

Statebuilding is slow, difficult, and uneven by nature. Traditionally, scholars have focused on the projection of state power from urban cores into far-flung, sparsely populated peripheries (e.g. Tilly 1990; Herbst 2014). Today, however, low-income and informal urban neighborhoods are often the most prominent zones of state weakness. Many are partially governed by criminal and other armed actors who compete with the state for residents' loyalty (Arias 2017; Lessing et al. 2019; Lessing 2020; Melnikov et al. 2020).

This within-city variation in state penetration is in some ways puzzling. Basic order, dispute resolution, and other governance should be easier to provide in dense, central areas. Reducing crime and violence are also straightforward ways to raise residents' perception of state legitimacy. Even small investments in long-neglected zones could make a big difference. Urban peripheries should be the low-hanging fruit of statebuilding (Herbst 2006).

Why then do so many governments fail to establish Weberian monopolies in their urban cores? This study suggests one answer: statebuilding efforts could exhibit "increasing returns:" low or even negative in areas with initially little state presence and accountability, and higher in areas that begin with moderate government capacity. If true, this would give bureaucrats and politicians disincentives to invest in marginalized neighborhoods, driving a path-dependent "neglect trap."

This hypothesis flows from the unexpected results of a city-wide statebuilding experiment in Medellín, Colombia—to our knowledge, the first of its kind. The city chose a small number of lower-income neighborhoods and tried to dramatically intensify a wide array of everyday safety, problem-solving, and other civilian municipal services. The experiment sought to answer three questions. First, could expanding non-coercive state presence and services improve citizen perceptions of state legitimacy? Second, could it reduce insecurity and street disorder? Finally, if successful, would this also crowd out the governance and perceived legitimacy of the city's ubiquitous neighborhood gangs?

At a minimum, we expected the answer to the first two questions to be "yes," with stronger effects where relative state presence began lower. Instead, average impacts were null on average, with heterogeneity running in the other direction: large, significant effects on security and legitimacy in areas with above median initial state governance, balanced out by mildly negative effects in sectors below the median. Meanwhile, perceptions of gangs did not change at all. These results revised our priors about the returns to statebuilding, generating new hypotheses about the conditions in which high returns are possible.

Driving the intervention's design and our initial predictions was a literature suggesting that residents' perception of state legitimacy is rooted in effective service-provision (Arjona 2016; Beath et al. 2020; Berman et al. 2011, 2013; Carter 2013; Gurr 1971; Horowitz 2000; Krasner and Risse 2014; Levi et al. 2009; Risse 2011). We also hypothesized that municipal services and problem-solving could improve security and order on the streets, whether directly by deterring or solving neighborhood problems before they escalate, or indirectly as improved state legitimacy increased residents' cooperation with authorities (Akerlof and Yellen 1994; Risse and Stollenwerk 2018; Acemoglu et al. 2020).

This civilian-led intervention contrasts with more police-focused approaches to improving legitimacy and security. Policing clearly has an important role to play, and research shows that increased police and military presence tends to improve security (Chalfin and McCrary 2017; Blair et al. 2019; Collazos et al. 2021; Karim 2020; Magaloni et al. 2020; Blair and Weintraub 2021). At the same time, there is a growing awareness that coercive approaches can backfire—marginalizing victimized groups, undermining state legitimacy, and even enhancing the reputation of nonstate actors (Cruz 2011; Owens 2019; Acemoglu et al. 2020; Gonzalez and Komisarow 2020; Lake 2022; Morris and Shoub 2024). Thus, many cities are looking for complements and alternatives to traditional security strategies.

The intervention we study was guided by the concept of *convivencia*, or "coexistence." The approach is common across Latin America, especially Colombia, where mayors have been experimenting with non-police security strategies for decades. Convivencia efforts often involve street-level bureaucrats and municipal agencies working to improve everyday order and state–community relations. This includes: fostering better communication between communities and authorities; strengthening community organizations and their ability to solve local problems; and connecting residents with state agencies that can solve more serious household and neighborhood problems. These activities parallel the way that local governments worldwide manage neighborhood problems through a mix of municipal agencies, social work organizations, and constituent-service offices. There is little evidence on such urban statebuilding efforts, however, and even less on the role of non-police efforts.[1]

We worked with Medellín's mayor and his government—the *Alcaldía*—to dramatically intensify regular convivencia activities in 40 low- and middle-income residential neighborhoods. To evaluate this, we and the city identified an experimental sample of 80 "sectors"—small, informal sub-neighborhoods of up to 10 city blocks with about 1,000–3,000 residents each. *Operación Convivencia* began in early 2018 and lasted 20 months. It represented a massive increase in central and street-level state attention in treated sectors.

We prespecified two primary outcomes: *Relative state legitimacy* and *Relative state governance*. These measures come from an endline survey of roughly 2,400 residents and businesses in the 80 experimental sectors. Legitimacy captured perceived trust, satisfaction and fairness of the state versus the gang, and governance captured the perceived involvement of each actor in addressing local security issues and disputes. We also explore impacts on objective measures of security, collecting administrative data on *Security-related emergency calls* and *Reported crimes*. Finally, we conducted hundreds of qualitative interviews before, during, and after the intervention, and surveyed the intervention implementers to assess compliance.

Despite the intensity of the intervention, we see no impact on our primary outcomes—at

[1] One experiment, in rural Pakistan, found that people told about improvements in state courts said they were more willing to use them, and were more willing to make cash transfers to the state (Acemoglu et al. 2020). Such information experiments are valuable, but they may be difficult to translate into real and sustained improvements in governance. Another experiment in Liberian towns focused on improving community dispute resolution practices and norms through large-scale training. The program reduced low-scale violence such as threats, fights, and property destruction over at least three years (Blattman et al. 2014; Hartman et al. 2021). While promising, this was a non-governmental rather than a statebuilding intervention, so the potential effects on state legitimacy are unclear. It is also unclear whether such approaches can work in large cities, and in the presence of criminal governance. Our intervention, randomized at the neighborhood level, is to our knowledge the first of its kind.

least on average. A huge increase in state attention did not raise trust and satisfaction with the state or improve perceptions of state responsiveness to insecurity; there was even a small, nonsignificant decline in state governance. On average, residents in treated neighborhoods did not register an increase in state increased presence.

These findings, if generally true, would challenge some of the current enthusiasm for civilian alternatives to policing, as well as the broader statebuilding strategy of increased service provision. Particularly so given Medellín's reputation as having one of the most capable municipal governments in Latin America, with ample fiscal resources (Leyva 2010).

A closer look, however, reveals unexpected heterogeneous effects, with nuanced implications for statebuilding. Prior to the intervention, some sectors received more attention and services from the Alcaldía than others. Many also had local gangs that provided some degree of governance. Anticipating that intensifying state attention could have greater effects where initial levels were lower, with diminishing marginal returns where the state was already present, we prespecified a subgroup analysis based on local leaders' assessments of initial state versus gang governance.

We indeed find heterogeneous effects, but in the opposite direction of our priors. In the subgroup with high baseline relative state governance ("high-state" for short), residents registered a large increase in services and attention. The program raised perceptions of state legitimacy by roughly 10 percent, while reported crimes and emergency calls related to fights and public disorder decreased by 40 percent. An index of all outcomes improves by a huge margin—0.67 standard deviations. In the "low-state" subgroup, however, residents did not register any improvement in attention. Perceptions of the state weakly worsened, and there was no change in reported crime or street-level disorder.

As for gang reactions, even where the intervention worked, we see no evidence that criminal governance changed. Our interviews suggest an explanation: gangs did not find municipal employees threatening, and may have indirectly benefited from civilian-driven order. Criminal groups in Medellín are mainly concerned with police presence, mainly to protect drug rents (Blattman et al. 2024).

The likeliest cause of these heterogeneous treatment effects was uneven implementation quality. Our data suggest that central agencies and street-level bureaucrats systematically fell short in initially low-state areas. Residents in these sectors did not notice more municipal personnel and activities on their streets, suggesting either lower attention from the staff or lower community engagement. Furthermore, the street-level staff assigned to those sectors reported more failures of the central task force and agencies to deliver on commitments. In contrast, in high-state sectors, residents noticed for more municipal personnel and activity in their neighborhood, and street-level staff seldom reported central failures.

These results suggest that intensified service provision *can* increase state legitimacy and improve local security—provided the state actually delivers. They also suggest that the state's initial penetration and engagement in a community shape its capacity and incentives to deliver there. Taken together, this would imply higher returns to investment in places where the state is already present, and low or even negative in more neglected areas.

Such increasing returns could arise for several reasons. One is that there may be start-up costs or entry barriers to statebuilding, such that establishing robust state governance and legitimacy requires large and sustained investments, especially from a low starting point. Alternatively, low initial state governance may be endogenous to a community's political power and other unobserved characteristics. Neighborhoods that are marginalized, disorganized, or have low social capital may have difficulty holding state agents accountable, leading those officers to devote more attention to the communities with greater influence and voice.

Whatever the source, increasing returns in statebuilding would create the conditions for "neglect traps"—a political analog to the development traps that can arise from increasing returns to investment (e.g. Duflo and Banerjee 2011; Weil 2008). Politicians and bureaucrats need to achieve observable results with limited resources in short time frames. If their returns are highest in well-governed areas, they will have little incentive to address disparities. Such neglect traps would help explain a common feature of cities in middle-income countries worldwide: zones of wealth and security abutting zones of relative state absence, disorder, and criminal governance. While speculative, Medellín's experience suggests this is an important hypothesis for future experimentation and research.

2 Context

2.1 The state and security in Medellín

Medellín is Colombia's second-largest city, with a population of roughly 2.5 million. It is one of the nation's industrial and commercial centers, with an annual income of roughly $11,500 per capita in purchasing parity terms. It has a well-organized bureaucracy with high tax revenues and public services.

The metropolitan police force has roughly 2.7 officers per 1,000 people—slightly higher than the U.S. national average, comparable to Los Angeles. Medellín is divided into 16 *comunas*, each one generally with its own police jurisdiction with a commander and station. Each station is also divided into a large number of *cuadrantes* (quadrants). Each quadrant has 6 assigned officers who patrol on motorbikes, in pairs, in 3 shifts per day.

In Colombia, however, the police are a national institution—a branch of the Defense Ministry. Although the constitution designates mayors as local police authorities, this only gives them influence over tactics and broad policy. The number of officers, their wages, and training decisions are made by the central government, not mayors.

Police autonomy is one reason why Colombian cities have been experimenting with civilian security measures for decades. Most cities have a large municipal agency, the Secretariat of Security, that directs a diverse array of activities and staff. Medellín's secretariat has roughly 1 staff per 1,000 residents, giving it roughly a third as many personnel as the police. Its budget and staff have grown enormously over time, from spending as little as 2 USD per capita in 1985, to 40 in 2007, and well over 50 in recent years (see Appendix A).

The mayor or *Alcalde* oversees the Secretariat and appoints all leadership positions. These leaders, along with other permanent and non-politically appointed senior staff, constitute the top-down task force that was part of the intervention we evaluate in this paper.

Secretariat staff are typically civilian professionals who specialize in social work, dispute resolution, family services, community outreach, or central administration. They provide a range of services to residents, including responding to various emergencies and street disorder, resolving disputes and domestic violence, or regulating the use of public space.[2]

These security- and dispute-related units have several headquarters per comuna, including *inspecciones* which directly resolve community disputes through a formal, fast-track justice service, and *comisarías* which provide a wide range of family services aimed at resolving legal problems, mental health problems, domestic violence, child protection, and family law.

The Alcaldía also has several mechanisms for receiving and responding to community complaints, such as streetlights, graffiti, or broken playgrounds. One is a citizen service hotline, called *Linea 123*. Another is a meeting called a *Consejo de Convivencia*, held once annually per comuna, where community members and senior city and police officials decide on a plan of action for tackling specific community grievances and issues. Finally, each comuna has one outreach staff member called a "liaison." These liaisons are tasked with community outreach, facilitating the Consejo, and linking citizens to the Secretariat's services.

2.2 Street gangs

In Medellín, gangs and criminal governance are also important features of everyday life. Virtually every low- and middle-income neighborhood has a local drug-selling gang called a *combo*. There are roughly 400 in the metropolitan area. Combo territories—often no more than 10 to 25 blocks—are well demarcated, known to residents, and have been largely stable since at least the early 2010s.

[2]In addition to these *comuna*-based city services, comunas are divided into neighborhoods called *barrios*. Each barrio has an elected community action board (*Juntas de Acción Comunal*, or JACs) that helps local groups regulate and organize their community. JACs are rarely involved in security, protection, and dispute resolution, however, and so they are not a major focus of the intervention or activities we study in this paper.

Since 2016 we have conducted semi-structured qualitative interviews with 149 gang leaders and members across 79 criminal groups. Obviously, this is a convenience sample of criminal actors who agreed to speak. Almost half took place in one of Medellín's three major prisons, from which leaders commonly direct street operations. Appendix B discusses human subjects protections and how we dealt with ethical issues associated with these interviews.

We found that Medellín's combos are generally small, well-organized, illicit firms whose profits come primarily from local drug retailing, supplemented by protection fees (described below). Combos typically have 15 to 50 salaried members, most aged 16 to 35, and each member typically has a well-defined position in one of the combo's illicit business lines. Besides holding a local monopoly on drug retailing and protection, combo members frequently participate in and regulate local informal and sometimes legal markets, including microfinance (loansharking) and consumer staples—especially cooking gas, arepas, milk, and eggs. Some also sell private protection services and other governance services in return for fees.

2.3 State and gang governance and legitimacy

There is virtually no systematic data on the governance activities or legitimacy of city governments, let alone gangs. We developed several new measures to assess their performance.

2.3.1 Design and measurement

In developing our survey measures, we drew on our interviews of criminal group members and additional qualitative interviews with 23 community leaders, 151 residents and shopkeepers, as well as 19 police officers and officials, 17 city officials, 10 prosecutors, and 18 other crime and security experts. These, together with our pilot surveys, revealed that residents and businesses were knowledgeable and open to speaking in private about their attitudes towards the police, Alcaldía, and local gang; the types and quality of services provided by each actor; and the taxes and fees each collected. (We investigate measurement error in more detail below and find little evidence of systematic misreporting.)

In December 2019 we conducted a representative survey of Medellín's 223 low- and middle-income barrios (excluding 27 high-income and non-residential barrios). The survey had several goals: to develop and test these new measures of state and gang governance and legitimacy; to inform theory and policy; to evaluate the impacts of historical natural experiments; and to serve as an endline survey for this experiment.

To create a sample representative of Medellín's low- and middle-income neighborhoods, we randomly sampled 2,300 blocks, stratified by barrio, and tried to interview roughly two households and one business per block (for an average of 21 respondents per barrio). In addition, for the experimental sample, we surveyed an average of almost 6 blocks in each sector eligible for the intervention, and tried to interview four households and one business per block (for an average of 29 surveys per sector). The representative and experimental samples are distinct. This section focuses on the representative city sample alone.

Each survey lasted roughly 30 minutes and was delivered in person by enumerators on handheld tablets. Enumerators were employed by one of the country's largest survey firms, and had no affiliation or identification with the government or the intervention.

Measuring legitimacy To measure residents' perceived *Legitimacy* of the state (police and Alcaldía) and combos, we took a bottom-up, empirical approach. Through semi-structured interviews, we explored the words, phrases, and concepts that residents used to describe organizations they trusted and expected to rule. Then we distilled these into simple questions that worked reliably on a large-scale survey instrument.[3]

A key insight from this qualitative research—borne out in our survey results—is that legitimacy is not necessarily zero-sum: combo legitimacy need not come at the expense of police or municipal legitimacy. We structured the survey to allow such variation, asking households about each actor separately: how much residents trusted each actor; whether

[3]Governance and legitimacy are both contested concepts, in part because they carry normative connotations (Collier et al. 2006). Legitimacy is particularly thorny, with some scholars advocating it be jettisoned altogether (e.g. Wedeen 2015). Our goal was to devise operational measures that could capture the key variation across Medellín's neighborhoods and also assess our theories and potential treatment effects.

Table 1: State and combo legitimacy and governance, barrio survey averages, 2019 (N=4,598)

	Frequency/Rate (0-1 Scale)				Relative State – Combo	
	State		Combo		City-wide survey	Experimental control group
	Estimate	SD	Estimate	SD		
	(1)	(2)	(3)	(4)	(5)	(6)
Legitimacy Index	0.58	0.21	0.43	0.28	0.13	0.13
When solving problems in the neighborhood:						
How much do you trust the...	0.57	0.30	0.36	0.36	0.19	0.20
How fair is the...	0.55	0.27	0.41	0.35	0.11	0.12
How do you rate the...	0.60	0.22	0.51	0.28	0.09	0.09
How would your neighbors rate the ...	0.59	0.23	0.50	0.29	0.09	0.08
How much do your neighbors trust the...	0.57	0.28	0.47	0.36	0.09	0.06
Governance Index	0.41	0.26	0.34	0.29	0.07	0.07
How often they intervene when:						
HH: Someone is making noise	0.43	0.38	0.19	0.30	0.23	0.26
HH: Home improvements affect neighbors	0.41	0.38	0.25	0.34	0.16	0.14
HH: There is domestic violence	0.51	0.37	0.35	0.37	0.15	0.15
HH: Two drunks fight on the street	0.54	0.36	0.40	0.37	0.13	0.13
Biz: Someone disturbs a business	0.50	0.38	0.36	0.38	0.12	0.16
Biz: You have to react to a robbery	0.52	0.37	0.40	0.39	0.11	0.12
Biz: It is necessary to prevent a theft	0.45	0.37	0.38	0.39	0.07	0.08
Biz: Businesses in this sector are robbed	0.42	0.39	0.35	0.38	0.05	0.07
HH: People smoking marijuana near children	0.29	0.36	0.25	0.36	0.04	0.03
HH: A car or motorbike is stolen	0.46	0.37	0.43	0.38	0.04	-0.01
HH: Someone is threatening someone else	0.42	0.36	0.41	0.37	0.01	-0.01
HH: You have to react to a robbery	0.46	0.36	0.45	0.38	0.01	-0.02
HH: Someone is mugged on the street	0.39	0.36	0.41	0.38	-0.01	-0.05
HH: It is necessary to prevent a theft	0.40	0.36	0.42	0.38	-0.03	-0.04
HH: Kids fight on the street	0.29	0.35	0.32	0.37	-0.04	-0.03
Biz: Someone does not want to pay a debt	0.17	0.31	0.23	0.35	-0.06	-0.05
HH: Someone refuses to pay a big debt	0.22	0.31	0.39	0.38	-0.16	-0.20

Notes: The legitimacy and governance indexes are averages of the component questions listed in this table. Columns 1–5 present averages from the city-wide survey, representative of Medellín's 223 low- and middle-income barrios, with 20–25 respondents per barrio. Column 6 reports averages for the experimental sample of 80 sectors, with roughly 30 respondents per sector. The Relative State measures in Columns 5 and 6 are the differences between columns 1 and 3. All legitimacy scales correspond to: 0 = Nothing, 0.33 = A little, 0.66 = Somewhat, 1 = Very. All governance scales correspond to: 0 = Never, 0.33 = Occasionally, 0.66 = Frequently, 1 = Always. Both households (HH) and businesses (Biz) were surveyed on governance levels (N=4,598), but only households were surveyed on legitimacy (N=2,950). For Column 6, the experimental sample, these sample sizes are 2,362 and 1,906.

each behaves fairly; how satisfied residents were with each; whether they thought their neighbors trust each; and how they thought their neighbors would rate each. Each question used a 4-item Likert scale, which we rescaled so that 0 = Nothing, 0.33 = A little, 0.66 = Somewhat, 1 = Very. Table 1 reports average responses, as well as a *Legitimacy index* for each actor that averages all five legitimacy measures (pooling the police and Alcaldía responses for the state measure). *Relative state legitimacy* is the simple difference between the state and combo averages. It ranges from -1 to 1, where positive values imply that the police and Alcaldía are more trusted, seen as fairer, and rated more highly than the combo.

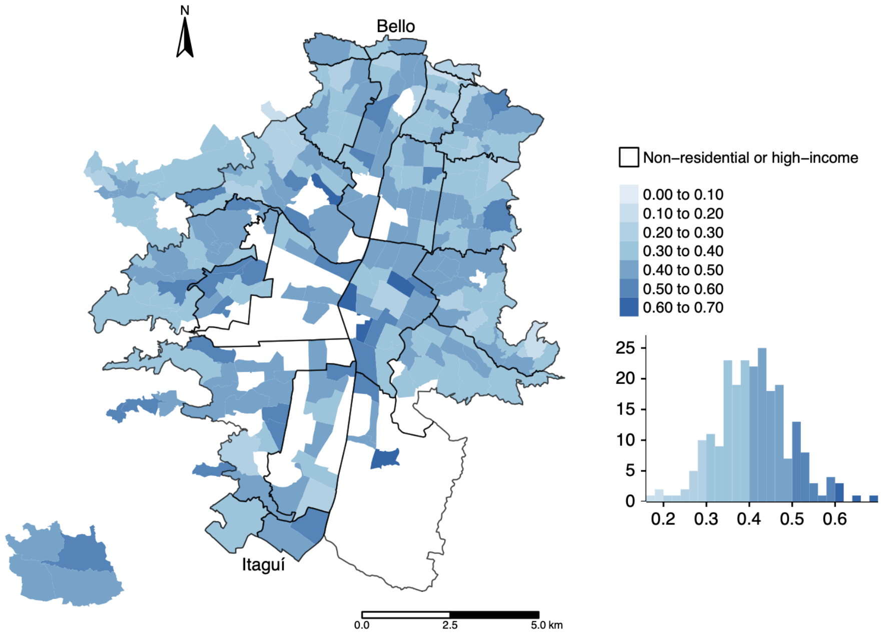
This focus on trust, satisfaction, and fairness follows other prominent measures that operationalize legitimacy using similar concepts (e.g. Levi et al. 2009; Karim 2020). This departs somewhat from the idea that legitimacy represents a population's willingness to obey and accept the actor's authority—the "right to rule" or the "license to govern" (Weber 1946; Tyler 2003, 2004; Gilley 2009; Risse 2011; Risse and Stollenwerk 2018). These dimensions of legitimacy are important, but we found that survey respondents struggled to understand these questions or articulate their attitudes to authorities in these terms.

Measuring governance To measure the core governance functions of order-provision and dispute resolution, the survey asked how often state and combo intervened in everyday forms of community disorder. We used our field interviews and observations to identify 17 situations that were both common and important to community members and businesses. These included neighbor disputes, street fights, debt collection, robberies, and other crimes. For each, we asked how often each actor intervened to deal with the problem—never, occasionally, frequently or always. As with legitimacy, we did not assume governance was zero-sum, and so asked separate questions for state and gangs.[4]

Governance has many domains, of course. Our measure focused on order-provision and dispute resolution for three reasons. First, they are some of the most foundational functions

[4]Unlike the legitimacy section, governance questions pooled the police and Alcaldía into one entity—the state—to reduce survey length.

Figure 1: State governance levels by barrio (N=4,598)



Notes: The figure displays average levels of state governance reported in each low- and middle-income barrio, using the average of all 17 items from Table 1, averaging across all survey respondents in the barrio. Darker shades imply higher responsiveness by the police and Alcaldía. We did not survey high-income residential neighborhoods or non-residential areas (in white).

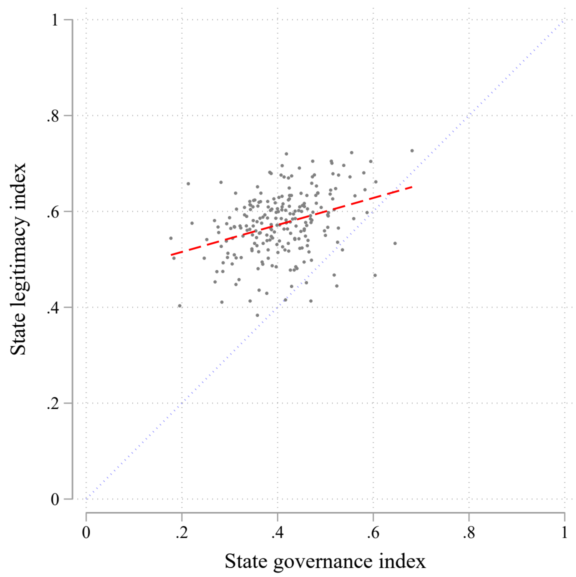
of the state. Second, these are the main areas where both the state and the gang provide services. Third, we wanted to be able to test whether intensifying civilian municipal services could result in greater perceived involvement of the state in basic order provision.

Table 1 reports average reponses. Each question used a 1–4 Likert scale, rescaled so that: 0 = Never, 0.33 = Occasionally, 0.66 = Frequently, and 1 = Always. We averaged these 17 items into 0–1 *State* and *Combo governance indexes*. We also calculate a *Relative state governance* measure ranging from -1 to 1.

2.3.2 State performance

When it comes to trust, fairness, and satisfaction, the average low- and middle-income resident has moderately positive views of the police and Alcaldía. The average legitimacy index is 0.58—a level that corresponds being "somewhat" fair or trustworthy on the Likert scale. Meanwhile, when it comes to responsiveness to neighborhood security and disorder,

Figure 2: Relationship between state governance and legitimacy by barrio, 2019



Notes: Each dot is a barrio average of household responses, and the dashed line indicates fitted values.

residents score the state as 0.41 on the 0 to 1 scale—slightly better than “a little” responsive to disorder and disputes. State responsiveness is greatest (above 0.5) for robberies, domestic abuse, and adult street fights (Table 1). State responsiveness is poorest (below 0.3) for debt collection, teenage disputes, and drugs and smoking near children.

As Figures 1 and 2 illustrate, however, there is widespread variation in perceptions of state responsiveness and legitimacy. Figure 1 maps state governance by barrio. Levels range from 0.2 to over 0.6. State legitimacy is less variable—barrio averages seldom fall below 0.4 and can be as high as roughly 0.75 (corresponding to slightly above 3 on the 4-item Likert scale). Not surprisingly, *State governance* and *Legitimacy* are strongly correlated, as seen in the barrio-level scatter plot in Figure 2. A regression of *State governance* and *Legitimacy* yields a coefficient of 0.28 ($p < 0.01$, with standard errors clustered at the barrio level).

2.3.3 Gang performance

Criminal governance is pervasive across the Americas (Uribe et al. 2022), and Medellín is a well-known case (e.g. Arias 2017; Cruz and Durán-Martínez 2016; Moncada et al. 2018).

After the state, combos are the most common organization that residents turn to in order to settle household and business disputes, collect debts, stop fights, prevent thefts, manage the homeless and drug addicts, and other neighborhood disorder. In return for these services, combos typically collect fees from local businesses and residents. This tax is sometimes called a *pago por la vigilancia* (surveillance fee) or, more colloquially, a *vacuna* (vaccine).[5]

As Table 1 illustrates, average combo governance is 0.34—about 83% the level of the state. Citywide, among the 17 components of this index, the combo is rated as slightly more responsive than the state in five: responding to muggings, preventing theft, teenage street fights, and business and household debt collection. Average combo legitimacy is 0.43—about 75% of the state average.

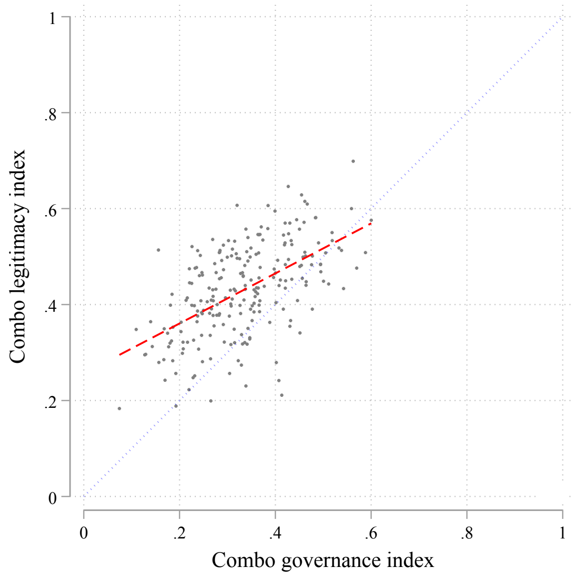
Governing helps combos win the trust and loyalty of residents, and as we can see in Figure 3, combo governance and legitimacy are positively correlated. But the horizontal spread in Figure 3 also illustrates how combos vary widely in the extent to which they provide governance and security. Every low- and middle-income neighborhood has a combo, but some combos provide little to no governance while others provide a wide range of services.

2.3.4 Relative state performance

Importantly, higher state governance and legitimacy do not necessarily imply less combo governance and legitimacy. Figure 4 illustrates the range of variation, and the low correlation between the two actors. Our interviews suggest several forces that work against the tendency of one actor to crowd out the other. First, to the extent demand for governance exceeds supply for both actors, more governance from one does not need to reduce use of the other. Second, both actors want to govern more where the willingness to pay is greatest. Finally, other research suggests that combos are primarily interested in protecting their drug rents, and respond to police presence by governing more (Blattman et al. 2024). Our theoretical

[5]In most cases, this is not pure extortion, in the sense of demanding money in exchange for not inflicting harm. Nonetheless, payment is seldom entirely voluntary. If the local combo decides to provide security services on a block, most shop-owners will be compelled to pay the vacuna.

Figure 3: Relationship between combo governance and legitimacy by barrio, 2019



Notes: Each dot is a barrio average of household responses, and the dashed line indicates fitted values.

section and model, below and in the appendix, capture these competing forces.

The result is a patchwork of "duopolies of violence," where both gangs and the state act as governing authorities to various degrees (Skaperdas and Syropoulos 1996; Lessing 2020). Figure 5 maps this variation, plotting *Relative state governance* by barrio. The state is rated as more responsive than the combo in most neighborhoods, though to varying degrees. In 31 percent of neighborhoods, the combo is rated as more responsive to disorder and disputes.

3 Intervention

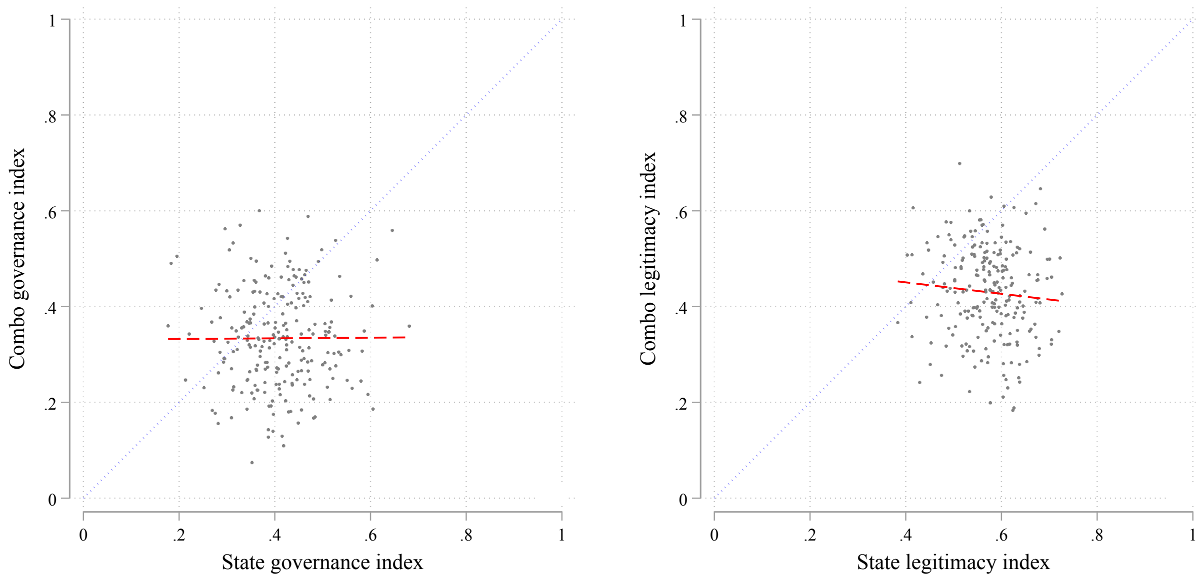
3.1 Background

Like many city governments, Medellín's Alcaldía was interested in statebuilding in the sense of increasing public safety, improving community relationships, and increasing its perceived legitimacy. Ideally, citizens would also seek out the government for security and dispute resolution rather than local gangs.

Given that the police are a branch of the national government, however, Colombian

Figure 4: Relationship between state and combo by barrio, 2019

(a) State-combo governance (b) State-combo legitimacy



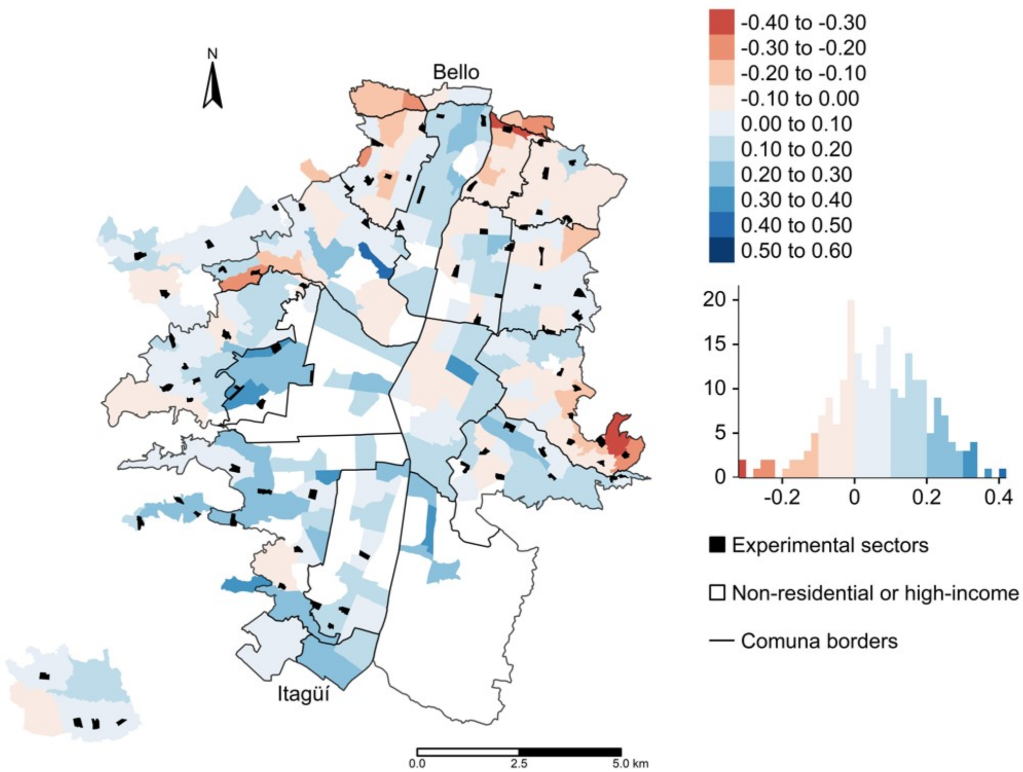
Notes: Each dot is a barrio average of household responses, and the dashed line indicates fitted values.

mayors have limited control over the number or capacity of the metropolitan police force. Thus, they must turn to other mechanisms of government. There are many civilian-led security approaches, from alternative dispute resolution, to community violence interruption, to increased social services—many of which exist in Medellín. In this case, the Alcaldía was interested in understanding the returns to intensifying the broad array of everyday municipal services and improving the functioning and governance of small neighborhoods.

There were two precedents for *Operación Convivencia*. One was the broad base of existing services offered by various municipal agencies, especially those under the Secretariat of Security. As discussed above, this included a range of dispute-resolution and family-services offices, a citizen-service hotline, each comuna's annual *Consejo de Convivencia* with senior city and police officials, and the comuna's existing liaison charged with facilitating community awareness of and access to these services.

The intervention was also inspired by a small and little-known governance and public-security effort in one of Medellín's poorer and under-served neighborhoods, La Loma. A

Figure 5: Relative state-combo governance and legitimacy levels by barrio, and location of experimental sectors (N=4,598)



Notes: The figure displays relative state-combo governance in every low- and middle-income barrio, averaging all 17 items from Table 1, averaging across all respondents in the barrio. Red-hued neighborhoods indicate that the combo is a net higher responder to disputes and disorder, blue-hued indicates the state is predominant, where in both cases darker hues are associated with greater predominance.

small unit in the Alcaldía assigned 7 liaisons to the neighborhood—about 1 per 23 blocks, a huge increase over the normal level of 1 liaison per 540 blocks. From 2012–17, these staff set out to improve legitimacy and governance by: (i) helping community groups better organize themselves to address local problems (such as neighborhood cleanliness, idle youth, or conflict resolution); (ii) connecting residents in need to existing city services (such as dispute resolution and mediation, or family and mental health services); (iii) bringing neighborhood problems to the attention of city agencies (such as garbage collection, broken lights, and playground repairs); and (iv) generally improving communication among the community, local government, and the police. Our qualitative interviews and observations in La Loma suggested that the intensification of street-level staff significantly increased community organization and access to municipal services, and the legitimacy of the state rose.

We brought these efforts to the attention of the Mayor and Secretary of Security and proposed that they test the returns to enhanced non-police state presence by experimentally intensifying both general service provision and liaison activity. They agreed.

3.2 Sample

Working with the Alcaldía, we determined that we could best evaluate the returns to city services and personnel by deploying a highly-intensive intervention in 40 small areas called "sectors." Each sector is an informal but well-defined neighborhood, far smaller than an official barrio. The government identified 80 sectors as an experimental sample, all drawn from the city's low- and middle-income barrios. They ranged in size from 200 to 600 households (1,000–3,000 residents), typically covering no more than 10 medium-density blocks.

The intervention was focused in small sectors of up to 10 blocks not because the approach demanded such intensity, nor because this change was the most important margin to evaluate. Rather, intervening intensively in a small number of sectors had several advantages. First, keeping the sectors small would help to maximize statistical power by increasing treatment intensity. Second, we wanted to minimize the chance of interference (e.g., spillovers) between treatment and control sectors. We worked with the city to ensure the sectors were at least 250 meters distant from one another. This limited the number of potential experimental units in the city. Finally, there were budgetary constraints. The Secretary of Security estimated he had funds to hire an additional 40 liaisons. One liaison per sector was the most straightforward way for the city to organize the new personnel.

Figure 5 displays the 80 experimental sectors. They were broadly representative of the city's neighborhoods in terms of their demographics, geographic features, and variation of state and combo governance. Columns 5 and 6 of Table 1 compare governance and legitimacy levels in the city and experimental samples, and we see that they are similar.

3.3 Intervention activities

Operación Convivencia began in April 2018 and ran for 20 months, until December 2019 when the mayor's term ended. The city designed the intervention to be delivered similarly across all sectors.[6] In each sector, the intervention had three main components:

Central task force The Alcaldía first created an inter-agency task force to respond to local concerns. This could include normal services—e.g., poor trash pickup or broken playground equipment—but the task force also tried to respond to security concerns, including attention from the city's dispute-resolution officers and family services. Concerns reached the task force via liaisons, other city staff, and also the general citizen hotline *Linea 123*.

Community–Alcaldía events The city also sought to improve communications and relationships with sector residents. First, officials from the Alcaldía and local police commanders were asked to attend twice-annual *Consejos de Convivencia*. These councils were established

[6]The Secretariat of Security's estimates for the total cost of the program over the 20 months of implementation are about COP 4,400 million or USD 1.1 million at market exchange rates (figures are for 2024). That is, about COP 110 million or USD 27.5 thousand per sector. The main costs correspond to the 40 full-time liaisons—which account for almost 90% of total costs—the field coordination team that oversaw liaisons' activities, staff training, and the targeting activities that preceded the implementation of the program. Proportionally speaking, this increase in staff and funding was large, roughly matching the increase in the level of investment in the civilian Secretariat of Security over the past 2–3 decades (see Appendix A).

by legal decree in the 1990s to increase the accountability of public officials to community problems. During these meetings, officials and community members identify specific problems and agree on mutual responsibilities and commitments. Normally, there is one consejo per comuna per year. Given that there are no more than 3,000 residents per sector, and about 150,000 people per comuna, a sector-specific consejo represents at least a 50-fold increase in communications and opportunities to address problems.

In addition to these meetings, the Mayor's office also organized a large one-time event called *Caravana de la Convivencia*: a weekend-long street festival in each sector where, in addition to music, food, and entertainment, representatives from each agency were on hand to explain their services in detail and identify residents in need of assistance.

Street-level liaisons The city also assigned a full-time street-level bureaucrat—a liaison—to each treated sector. Normally, the city has one liaison for each of the 16 comunas—roughly 1 per 540 blocks. For this intervention, the city hired 40 new liaisons as contractors. Thus treated sectors had 1 liaison per 9 blocks—a roughly 60-fold increase in street-level staffing.

Liaisons were expected to spend 3–6 days per week in their assigned sector, and otherwise work in the Alcaldía offices. They were given a high level of autonomy to engage and mobilize the sector as they saw fit. Still, liaisons had weekly targets and quotas for neighborhood events and resident referrals. Unfortunately from a research perspective, only major liaison activities and task force responses were formally logged and geolocated (about 10 per month). This was, however, by design. The liaisons—relatively young, new hires—were enthusiastic, active, and motivated to work in their sectors, and the Secretariat was concerned that too much paperwork would crowd out their time in the field or constrain their actions and decision-making. Liaisons had multiple roles, including:

- Collect and formally register community concerns to the inter-agency task force in the Alcaldía, and lobby to see that these concerns are addressed.

- Organize community events and meetings, including but not limited to the *Consejos* *de Convivencia* and *Caravana de la Convivencia* mentioned above.

- Help community organizations coordinate local collective action (e.g., deciding garbage spots and dog excrement norms).

- Provide training to community leaders and organizations in dispute resolution and related skills, and encourage them to take an active role in resolving local issues.

- Proactively identify individual and neighborhood problems and refer them to the relevant city agency for assistance (e.g., connecting residents with interpersonal conflicts to the comuna's *inspecciones* for dispute resolution or *comisarías* for family problems).

- Work with police officers to better inform community members of the "police code"—the country's legal guidelines for dealing with and correctly reporting nuisances, misdemeanors, and crimes, what officers were permitted to do, when to call them, and when to approach the Security Secretariat.

Like roughly 70 percent of all municipal staff, liaisons were employed on a contract basis through a non-governmental organization with extensive experience providing neighborhood outreach. They had a manager in the Security Secretariat that trained them, monitored their activities, and controlled quality.

Liaisons were not residents of their assigned sector, though they all came from low- and middle-income communities in Medellín, most of which would have had a combo and a degree of criminal governance. Rather, liaisons were professional staff hired for the position, with profiles similar to the city's existing cadre of liaisons: university-educated (often in the social sciences, psychology, or social work), ages 25–35, and about half men and half women.

3.4 Activities in control sectors and by police

The Alcaldía took steps to minimize changes in services to other neighborhoods (including control sectors). The 40 treatment liaisons were newly contracted for the intervention, leaving existing liaison staff unaffected. As for top-down attention, although the city did not increase centralized city staffing, treated sectors represent just 2.5 percent of Medellín's blocks. Any reduction of services received by the other 97.5 percent (including treatment sectors) should be minimal, and very unlikely to affect the validity of estimated treatment effects. We use the 2019 representative survey to test for this and, as described below, rule out significant changes in service provision nearby.[7]

As for policing, the Alcaldía designed the intervention to minimize any direct change in policing levels across treatment and control sectors, and we observed no obvious changes in policing attention. It is possible that the intervention influenced police patrols indirectly, however. On the one hand, improving communication with the police could have increased their presence or accountability. On the other hand, to the extent the civilian activities improved order on the streets, police would have been been called less for emergencies.

4 Conceptual framework and predictions

4.1 Expected impacts on legitimacy and security

We had theoretical and empirical reasons to believe that *Operación Convivencia* could increase state legitimacy, as well as real and perceived efficacy at delivering security.

Legitimacy Several literatures connect the visibility and efficacy of state services with increased satisfaction, trust, and loyalty. For decades, scholars of rebellion, protest, and conflict have argued that when governments provide security and address grievances, beneficiaries reward authorities with support (Gurr 1971; Horowitz 2000). A similar principle underlies a great deal of foreign aid, especially in conflict zones, and there is suggestive evidence that development aid and security produce cooperation with the state (Berman et al.

[7] Appendix B details decision-making and ethical considerations, noting that to the extent that the program changed any service provision in unobserved ways, the program is within the range of normal variation in place-based municipal policies.

2011, 2013; Beath et al. 2020). Granted, states have many ways to build trust and the right to rule—including religious authority, ethnic appeal, participatory mechanisms, and procedural fairness (e.g. Tyler 2003; Risse and Stollenwerk 2018). But effective performance is a longstanding strategy for producing stability and popular support for the government (Levi et al. 2009; Carter 2013; Krasner and Risse 2014). This includes the building of relationships between citizens and street-level bureaucrats (Karim 2020).

Our qualitative work bolstered this view. In community interviews across the city, the speed and quality of services seemed to be the primary driver of confidence in the state. Our observations in La Loma suggested that residents rewarded the extra state attention with trust and collaboration. Our work during the early days of the intervention bore this out. For instance, as one liaison remarked, "Community members expressed things like: 'We have never been this close to anyone in authority before'," and went on to say that "They were very grateful for it. They welcomed us warmly into the community. It was an opportunity to show them different ways of doing things that they were completely unaware of."

Security The connection between Operación Convivencia's activities and neighborhood security is less obvious, because the 17 forms of crime, disputes, and disorder we measured include several in which the liaisons, the Consejos, and the central task force were unlikely to intervene directly. Generally these bureaucrats do not disrupt street fights, respond to thefts, or collect debts, though it is possible that their physical presence has a deterrent effect. The city's dispute-resolution officers certainly work to resolve conflicts, but they are not a rapid-response force reacting to everyday street disorder or violence.

Therefore, to the extent that the intervention might affect order and security, the most promising mechanisms are indirect. We hypothesized several channels. The first is through the actions of community organizations. The liaisons worked with local leaders and organizations to build their conflict-resolution and problem-solving skills. They also tried to shape collective beliefs about appropriate behaviors, as well as rules and fora for resolving disputes.

Similar skill acquisition and norm changes underlie most alternative dispute-resolution programs (Mnookin 1998), and seem to have played a major role in the success of one such program in Liberia (Blattman et al. 2014; Hartman et al. 2021).

A second possibility is that disorderly people avoid, or change their behavior in neighborhoods with more visible state presence or more active community organizations. Tackling minor problems and disorder could also avert escalation into larger and more violent disputes. This is consistent with evidence that urban renewal and interventions that enhance communities' ability to maintain public spaces contributed to reductions in crime and violence in the U.S. (Cassidy et al. 2014; Braga et al. 2015).

Third, the consejos and liaisons were intended to educate the public and manage expectations. For instance, town halls with municipal officials could improve communication, create realistic expectations, and increase perceived effectiveness and legitimacy of the government. Liaisons also taught residents about the "police code"—the responsibilities of officers and the limits on their authority. They also worked to improve communication between the community, the police, and municipal officials through town halls and community-awareness campaigns. "Some people didn't know what the 'Casa de Justicia' is," one liaison explained, "or what the 'Comisaria de Familia' does, or that there's the possibility of free conciliation in a Conciliation Center. So, when they learn about these services and use them, it generates more trust."

4.2 Expected impacts on gang rule

We also hypothesized that the intervention might reduce citizen dependence on combo governance. In part, we viewed the intervention through the lens of duopolistic competition, with the state and the combo offering residents distinct but substitutable governance services. In 2018, we did not have the benefit of the detailed data in Section 2 showing the correlation between state and combo governance is relatively low. Nor had any studies proposed that gang and state governance could be strategic complements when there are drug rents to protect, as Blattman et al. (2024) later proposed. Instead, we hypothesized that, should the state exogenously increase supply of its services, its relative "market share" should rise, while combo governance should fall. We illustrate this theoretical possibility with a simple formal model of imperfect competition in Appendix C.

This logic suggests that states can crowd out gang rule by improving the quality and reach of their services, echoing two related literatures. One attributes the emergence of organized crime and criminal governance to a power vacuum left by weak states (Gambetta 1996; Skaperdas 2001; Skarbek 2011). Meanwhile, scholars of rebel governance generally see the state and nonstate actors engaged in "competitive state-building," (Kalyvas 2006), providing security and protecting basic rights in order to earn popular support and legitimacy (Mampilly 2012; Staniland 2012; Arjona 2016; Blair and Kalmanovitz 2016). Our initial qualitative work bolstered this view. As one liaison remarked, "Some of these sectors were forgotten places, with no institutional presence. There were situations or issues that could be addressed, and the community realized that things could be done differently, because not everything can be handled by the combo."

Later, observing the intervention in progress, we moderated this view for two reasons. First, neither the state nor the combo appeared to fulfill community demand for governance. Citizens have a huge range of everyday disputes, neighborhood violence is commonplace, and neither the state nor the combo seem to respond to them all. The data later bore this out, as both state and combo governance measures are well below 0.5 in Table 1 above. Second, through 2018–19 we observed that combos generally regarded the liaisons and regular city services as benign. Combos' main concern was the police. While the intervention sought to increase police–community communication and understanding, there was no mandated change in police presence, and the effect on resident calls to the police was ex ante ambiguous.

4.3 Predicted heterogeneity

Finally, based on the literature and qualitative observations described above, we hypothesized that investments in state governance and capacity would be more effective in places where it began lower at baseline—provided the treatment was uniform. For this reason, we pre-specified a heterogeneity analysis based on initial state governance levels of sectors.

While our expectation was stronger effects in low-baseline sectors, we recognized that the direction of effects was hard to predict. In addition to the theoretically indeterminate response of combos, some research suggests that there is a risk that increasing state presence in a place where it has little previous reach could backfire. Many non-state actors—traditional authorities, community organizations, or local "big men"—govern in the absence of the state. These groups don't necessarily undermine the state, and are often complementary (Cammett and MacLean 2014; Van der Windt et al. 2019; Blattman et al. 2014; Henn 2021). When it comes to public security and justice, however, state legitimacy can suffer when other coercive actors—criminals, paramilitaries, or insurgents—govern the population (Berman and Laitin 2008; Acemoglu et al. 2020; Cammett and MacLean 2014).

Our advance qualitative work suggested that increased civilian state presence was likely to be met with enthusiasm from residents and community actors, and indifference from criminal actors, who are more concerned with police. Generally speaking, our results below bear this out. But it's important to note that in other contexts we might expect a different state–community relationship.

5 Empirical strategy and data

We preregistered our design, outcomes, estimation, and heterogeneity analysis, and report results with only minor deviations. Appendix D.1 contains details, and we enclose an anonymized version of the preanalysis plan.

5.1 Outcomes

In line with the above predictions, we prespecified two primary outcomes: Residents perceptions of *Relative state legitimacy* and *Relative state governance* in the area of security and order. To capture objective effects on security, we supplement these survey-based outcomes with administrative data on two exploratory outcomes that were not prespecified: reported crimes and calls to the city's emergency line.

Survey-based legitimacy and perceived governance Both outcomes are measured in the same manner as the city-wide survey described in Section 2 and Table 1. We ran the survey in December 2019, 20 months after the intervention began (during the final month of the intervention). We surveyed almost 6 blocks in each experimental sector, on average. We surveyed about four residents and one business per block, for a total of roughly 29 respondents per sector. Endline sample means for the 40 control sectors are reported in Column 6 of Table 1.

We made two important conceptual choices when designing these outcomes. First, we prespecified that we would look at impacts on *relative* performance between the state and the combo. In retrospect, absolute state legitimacy and perceived governance are arguably the more direct and appropriate primary outcomes for our core hypotheses. Thus, we report impacts on absolute state and combo outcomes, but treat them as exploratory.

Second, when it comes to measuring perceived governance, we intentionally focused on security and dispute-related governance. As discussed in Section 4, one of our principal research questions was whether a broad package of civilian-led activities can affect disorder and disputes—normally the focus of police. A broader governance measure could have included the maintenance of public space, the provision of public goods, the level of access and communication, as well as the quality of collective decision-making and coordination. Such data would enable us to directly measure the "first stage" effects of the intervention on the specific services offered by the program. We do examine several such indicators of delivery compliance, but our focus is on the "ultimate" outcomes we theorized.

Administrative data on security For crime, we create a *Sentence-weighted crime index* that aggregates all crimes reported in the 20-month intervention period within a 125-meter radius of each sector.[8] The index ranges from 0 to 1, with crimes weighted by their severity (proxied by sentence length guidelines for each crime). Note that these are formally reported crimes only. In Colombia, crimes can only be reported at a comuna's central police station, the local office of the Attorney General, or online for a specific set of crimes. Speaking to a police officer or calling the police will not result in a formal crime report (a fact that is widely known). Thus, reporting requires either traveling up to a kilometer to a station to fill out forms, or a long-form-completion process online. Evidence from Colombia's largest city, Bogotá, suggests that thefts of vehicles and other high-value items are frequently reported (for insurance purposes), as are crimes that result in serious injury or death. Because of the hassle, however, most petty crime goes unreported (Blattman et al. 2021).

We also count the total number of *Security-related emergency calls* to the city's hotline per sector. The vast majority of calls report a street fight, a case of domestic abuse, or a drug-related complaint—either a concern about a drug seller or (more commonly) drug users causing a public disturbance or loitering. Thus, these calls are primarily a measure of disorder as well as a measure of a person's likelihood of reaching out to the city. All calls are logged and geolocated to an address when the Secretariat of Security or police respond.[9] The coding of type of incident is relatively crude, however, and we can primarily distinguish between physical altercations (mixing domestic and street disputes), narcotics-related nuisances, and armed fights.

[8]We chose 125 meters for both measures because of our requirement that every sector be at least 250 meters from one another. 125 meters is half this distance, ensuring no overlap. Patterns are qualitatively similar for other radii. Note that there are no data on the frequency of police patrols by sector, so it is impossible to know the effects of the intervention on normal police presence outside of these demand-driven calls.

[9]Administrative logs say that more than 97 percent of calls receive a response and are geolocated. We do not know the location of the 3 percent of unanswered calls, so we cannot assess program impacts on police response. Since nonresponse is low, it seems unlikely to qualitatively affect the results.

Note that both measures are affected by the probability a resident reports a crime or calls. If the intervention reduced crime and disorder on the street, but also increased the likelihood that people collaborate with the state and report incidents when they do happen, then our estimates will understate the true improvement in order.

5.2 Randomization and estimation

To randomize, we used a matched-pair design, which shows balance along most covariates. Appendix D.2 describes the construction of matched pairs and presents tests of randomization balance.

We estimate intent-to-treat effects via the simple OLS regression:

$$Y_{isb} = \beta T_s + \gamma X_s + \alpha_b + \varepsilon_{isb}$$

where Y is the outcome from survey respondent i in sector s and matched pair b; T is an indicator for random assignment to treatment; X is a vector of the four main baseline indexes; and α_b is a vector of matched pair fixed effects.

There are three main threats to identification. The first is the moderate number of clusters—80, in 40 matched pairs. One long-recognized problem with clustered robust standard errors (CRSEs) is that they exhibit finite sample bias when the clusters are few in number (especially 30 or less) and tend to over-reject the null of no effect (MacKinnon et al. 2023). Our sample is well above the rule-of-thumb level, but for transparency and conservatism we report randomization inference (RI) p-values from 10,000 placebo treatments alongside the CSREs (Gerber and Green 2012).

A second threat is potential interference between units, such as spillovers. We designed the intensity and spread of experimental sectors to minimize this risk, and find no evidence of such interference, at least spatially. Using our representative city data, we see no evidence that the intervention affected blocks outside the experimental sectors, including control sectors (see Appendix D.3). We are unable to assess non-Euclidean spillovers, but note that the intervention was conducted in less than 2.5 percent of city blocks. If spillovers affected faraway rather than proximate blocks, these effects were probably thinly spread, since there is no reason they would concentrate in the 40 control sectors.

Finally, since our primary outcomes come from surveys, we also need to be concerned that citizens under-report gang activities, attenuating estimated treatment effects. Appendix D.4 discusses measurement error, and why it is unlikely to influence our results. Briefly, combos are a part of everyday life and not systematically stigmatized. We also designed a survey experiment and find no evidence of response bias.

5.3 Heterogeneity analysis and baseline data

Following the predictions outlined above, we prespecified heterogeneity analysis by baseline levels of *Relative state governance*.

Given the early stage of our research when the research opportunity arose and the short time before its implementation, we measured this differently than at the endline representative resident and business survey. We interviewed 80 local representatives of the Secretariat of Security and 149 resident leaders about the 80 sectors—roughly 3 individuals per sector. A few weeks before the intervention, we asked them "Who solves [problem] in this sector?" or "Who gives permission for [activity] in this sector?" for 14 different types of governance services. This included 9 order- and security-related items (such as dispute resolution, assaults, and thefts) and 5 other forms of governance (maintenance of and permission to use infrastructure, poverty relief, and permission for various political activities). Response options were "Mostly the combo", "Mostly the state and official community leaders", or "Both in proportion". Our relative state governance measure averages all 14 questions and rescales the index to 0-1, where a 1 would imply they replied mostly the state for all 14 items. This baseline measure has a mean of 0.405 and a standard deviation of 0.217. We look at impacts in two subgroups, above- and below-median initial relative state governance, via the OLS regression:

$$Y_{isb} = \beta T_s + \delta(T_s \times Low_{sb}) + \lambda Low_{sb} + \gamma X_s + \alpha_b + \varepsilon_{isb}$$

Variables are the same as in Equation 1. *Low* is an indicator for block-pairs with below-median relative state-combo governance.[10] To split the sample, we used the mean baseline relative governance index for the block-pair.

There are three limitations to this analysis. First, city and community leaders could have inaccurate or biased reports. A leader politically aligned with the municipal government might speak more favorably of the state's presence at baseline than the average sector household. Second, we do not have separate state and combo governance measures, meaning we cannot disentangle low initial state governance from high initial criminal governance.

Both of these are specific examples of the third and more general limitation of all heterogeneity analyses—they rest on an assumption of unconfoundedness. Is it relative state governance that drives differing results in the two subgroups, or is it some other sector characteristic that is simply correlated with relative state governance?

The data suggest that baseline relative state governance is relatively uncorrelated with most baseline variables, with the exception of other governance measures (Appendix D.5). Some of the strongest associations are with observed presence of the police and Alcaldía (as well as, in control sectors, endline absolute state governance). Lower baseline governance is not driven by higher criminal governance either. Hence the measure seems to be a reasonable proxy for initial state penetration. There are many other unobserved sector characteristics, however, and we return to alternative interpretations after examining the results.

[10] Given this interaction term, β now estimates the program impact on relatively high-state penetration neighborhoods, δ estimates the difference between high and low state penetration neighborhoods, and $\beta + \delta$ is the impact on low state penetration neighborhoods. Appendix Table D.2 shows that treatment-control balance within the subgroups. Note that the preanalysis plan specified that we will perform heterogeneity by initial level of criminal governance, but this was a misnomer, as we only ever had the relative measure available.

6 Results

We first analyze treatment compliance using several "first stage" proxies for program implementation, then turn to our main results.

6.1 Implementation quality and compliance

We collected four forms of quality and compliance data: qualitative observation of sector activities; administrative data on major activities logged by liaisons, endline survey data on residents' perceptions of state presence, and a post-intervention survey of liaisons. While we find little variation in major events logged, both resident perceptions and liaison reports suggest that day-to-day program activities and broader municipal engagement was poorer in places with low baseline relative state governance.

Qualitative observation Our research staff conducted spot visits and qualitative field observations during the first two months of the intervention. Overall our impression was one of resourceful, enthusiastic, hardworking efforts by skilled young professionals, and a good-faith effort by the city to implement the intervention. Our spot visits suggest that liaisons spent several days or evenings per week in their sector, held regular community events, and attempted to meet referral quotas, all within the few blocks they were assigned to. The Secretariat, consistent with the professionalism of the program, monitored liaison performance and replaced under-performers.[11] We did not observe any pattern between the quality of the liaisons and the types of neighborhoods where they were assigned.

Survey-based resident perceptions To assess whether treated residents noticed the increase in municipal government activity, our endline survey included six questions on Alcaldía personnel and activities in their neighborhood. In order to be relevant in both treatment and control sectors, these questions were deliberately general—whether residents saw municipal

[11] Within the first 8 months, the Secretariat had replaced half of the liaisons with more able personnel.

employees in the neighborhood, or knew about and attended events. Column 2 of Table 2 reports average treatment effects on each questions as well as a family index averaging all, to reduce the number of hypotheses tested.

On average, we see no evidence that residents noticed the increase in Alcaldía activity, or that they attended more events. The average change in the overall index is 0.01, less than 3 percent of the control mean. Only one of the six components is positive and statistically significant using RI p-values—seeing municipal staff in the sector, which rose roughly 8 percent relative to the control mean. Interacting with Alcaldía staff rose 11 percent relative to the control mean as well, and is significant using p-values from CRSEs but less so with RI p-values. These signs are promising, because everyday liaison street presence (rather than occasional events) is probably what citizens should have noticed most, and may be the best proxy for overall implementation.

However, our prespecified heterogeneity analysis—based on being above or below median initial relative state governance—reveals significant variation. Columns (3) and (4) report ITT estimates and randomization inference p-values for each subgroup, and Column 5 reports the difference. In high-state sectors, residents report a 16 percent increase in municipal activities and participation; in low-state sectors, they reported a roughly 12 percent decline. The divergence in the index between the two subgroups is dramatic—0.09, equivalent to 27 percent of the control mean. We see this divergence in every component of the index (Column 5). In high-state sectors, residents and businesses were dramatically more likely to notice and interact with municipal staff and be aware of and attend community events.[12]

Administrative data The only source of administrative data on compliance available are liaisons' logs of large-scale and long-term events and activities, roughly 10 per sector per

[12]Why might we observe a decrease in low-state sectors? The effects are concentrated in community meetings rather than staff visibility. One possibility is that, over 20-months, residents in poorly-governed sectors became aware that events were held, but too late to attend. Or they may have been less likely to attend meetings because they perceived as less useful. Moreover, the answers to these questions are inherently subjective, and may simply be capturing residents' sentiments towards the liaisons and activities—if expectations went unmet, there may be negative attitudes towards events.

Table 2: Did citizens notice and participate in increased state activity? Survey-based measures; average treatment effects and heterogeneity by initial relative state governance

Dependent variable	Control Mean (SD) (1)	ATE Estimate (CRSE p) [RI p] (2)	Het. by baseline rel. gov. Above median Estimate (CRSE p) [RI p] (3)	Het. by baseline rel. gov. Below median Estimate (CRSE p) [RI p] (4)	Het. by baseline rel. gov. Difference Estimate (CRSE p) [RI p] (5)	N
Index of first-stage variables (0-1)	0.33	0.010	0.052**	-0.038*	0.090***	1,908
	(0.27)	(0.395)	(0.000)	(0.009)	(0.000)	
		[0.289]	[0.016]	[0.058]	[0.005]	
Saw Alcaldia staff, binary	0.61	0.049**	0.090**	0.003	0.086*	1,892
	(0.49)	(0.011)	(0.004)	(0.849)	(0.014)	
		[0.044]	[0.031]	[0.456]	[0.062]	
Interacted with Alcaldia staff, binary	0.24	0.027	0.059*	-0.008	0.067*	1,892
	(0.43)	(0.083)	(0.028)	(0.587)	(0.038)	
		[0.147]	[0.090]	[0.367]	[0.088]	
Attended state events, binary	0.21	-0.002	0.033	-0.041	0.074*	1,876
	(0.41)	(0.876)	(0.079)	(0.031)	(0.006)	
		[0.460]	[0.135]	[0.119]	[0.050]	
Knew about state events, binary	0.52	0.031	0.106**	-0.054	0.159**	1,876
	(0.50)	(0.245)	(0.004)	(0.109)	(0.001)	
		[0.228]	[0.037]	[0.184]	[0.026]	
Attended community events, binary	0.10	-0.012	0.007	-0.034*	0.041	1,856
	(0.30)	(0.236)	(0.635)	(0.014)	(0.054)	
		[0.222]	[0.384]	[0.069]	[0.107]	
Knew about community events, binary	0.30	-0.037*	0.018	-0.097**	0.115**	1,856
	(0.46)	(0.054)	(0.331)	(0.002)	(0.001)	
		[0.083]	[0.274]	[0.012]	[0.012]	

<return>

Notes: This table reports answers to six Yes/No questions in the survey regarding whether residents and businesses noticed municipal employees and events or attended them. Each row is a different dependent variable. Column 1 reports control sector means. Column 2 reports ITT estimates. Columns (3) to (5) report treatment heterogeneity in sectors above and below the median level of baseline relative state governance, and the difference between the two groups. We report randomization inference p-values in brackets.

month. Liaisons were not required to log everyday facilitation, referrals, and other activities, including those of other agencies. This was to avoid paperwork burdens that might reduce liaison activity and autonomy in the field.

Most of the logged activities fit into 3 categories: organizing large public events; organizing consejos and meetings with police and other public officials and agencies; and helping to resolve major disputes.[13] In contrast to the resident surveys, liaison reports of major events do not vary by initial relative state governance. Panel (a) of Figure 6 plots logged activities against this heterogeneity measure. This suggests that liaisons' organization of major events was even across sectors, consistent with our qualitative observations and spot visits.[14]

If municipal staff were less active or visible in the neighborhoods with lower initial state governance, this could be because the liaisons performed fewer unlogged activities, or because other municipal staff did not serve the community adequately or up to expectations. The latter is consistent with our liaison surveys, which suggest that central state compliance was lower in the sectors with low initial relative state governance.

Liaison surveys Once the intervention ended, we interviewed all liaisons, asking a mix of open-ended and structured questions. We asked them to rate the central government's follow-through on promised service-delivery, on a 0–1 scale from full compliance to complete failure to deliver. Panel (b) of Figure 6 plots these responses against our main heterogeneity variable. On average, liaisons rated the Alcaldía's compliance at roughly 0.34, meaning the state "sometimes" failed to deliver on the requested support, but liaisons in sectors with relatively low state governance reported failures twice as often.

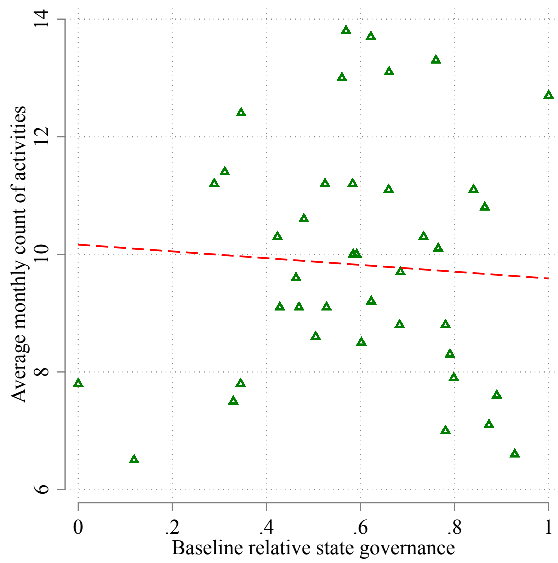
Our qualitative interviews with liaisons suggest that the most common problems were failures to deliver on community requests and meet expectations. For example, playgrounds

[13] Appendix D.6 maps activities by treatment and control sectors. Most of these major activities were held within a 125 meter radius of the sector, because community centers, meeting spots, and community organization offices were not always located in the sector itself, which had up to 10 blocks. Unfortunately the city does not maintain similar administrative data on the activities of the its regular 16 liaisons covering all comunas, but given their limited reach (1 per 540 blocks compared to 1 per 9 blocks in treated sectors) we presume that control sectors received no more than 1–2 percent as many events or activities.

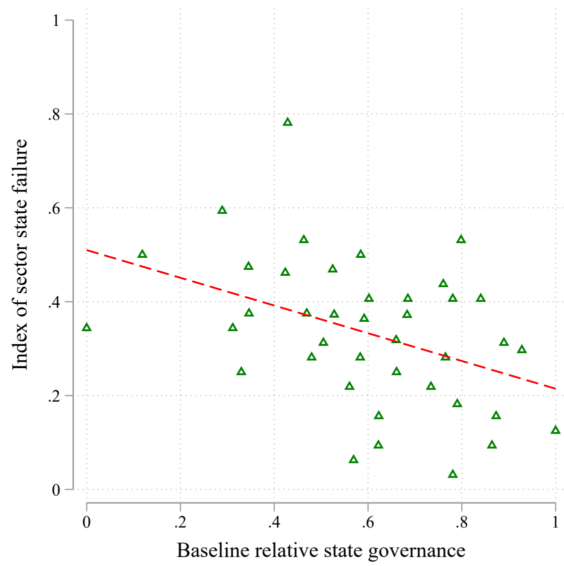
[14] Appendix D.6 confirms there is no difference between the subgroups.

Figure 6: How treatment experiences varied by initial levels of relative state governance (treated sectors only)

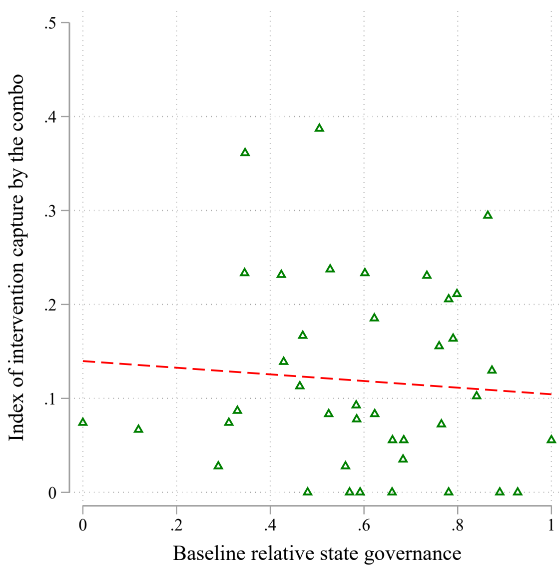
(a) *Count of treatment activities*



(b) *Alcaldía failure to deliver*



(c) *Instances of combo interference and capture*



Notes: The city required liaisons to log their activities, and Panel (a) reports the number of activities they logged, by levels of baseline relative state governance. Panels (b) and (c) contain data from a post-program survey of all liaisons. Based on their responses, we created two indexes. Panel (b) reports the frequency of various failures of the liaison or the wider state apparatus to deliver on promises. This includes a scale of the perceived frequency of failures and binary variables for whether specific local state agency failed. Values closer to 1 mean higher state failure. The data in Panel (c) capture the degree with which the combo interfered with liaison activities.

and public architecture often went unrepaired, despite requests. Or, as one liaison explained, "I managed to gather more than 60 people for the Consejo de Convivencia, but no one from the city showed up." Another liaison reported that the dispute-resolution officer "never came to [this sector] during all the time I was there. And he never gave us an answer to why he did not."

Several liaisons also reported facing difficulties due to low police quality or responsiveness. "The police have very little credibility," said one, "I had a police station near my territory and, honestly, I rarely saw patrols come in here." Another said how they had publicized the new police code—including official guidelines for when citizens should call the police versus civilian security and services agencies—but the residents were frustrated because the police did not follow it reliably. "It's very difficult to talk to people about the rules when they are witnessing a different behavior from police in practice," the liaison explained. Such responses suggest that heterogeneous "first-stage" survey results could be driven by prior experience of state governance rather than uneven treatment compliance.

Combo reactions Finally, we monitored combos' reactions to and potential interference with the intervention. Combos customarily monitor newcomers to the neighborhood. Therefore, as expected, almost all liaisons described having to explain their presence to the combo. However, we see no evidence that combo reactions shaped program impacts, with most liaisons reporting combo indifference to their activities. Two-thirds of liaisons reported no interference over the 20 months. The other third mostly said that the local combo was merely watchful, such as observing public events and meetings from a distance.[15] None of the liaisons reported extended harassment, violence, extortion, or bribes.

There is also no evidence of heterogeneous combo response, as illustrated in Panel (c) of Figure 6. The "combo capture" index aggregates several measures: a scale for the frequency

[15]There were exceptions, but this mostly affected the first few weeks of the intervention. For example, in two sectors, the combo initially prevented two liaisons from entering into the community for the first 2–3 weeks, but once the liaisons were able to explain their job and role, they were permitted to enter and perform their jobs without interference.

and difficulties of interaction with local gangs; an indicator for whether the gang ever took credit for the intervention; and a set of binary variables for activities by which the gang helped the liaison. Values closer to 1 represent higher involvement from the gang. In general, the average level is low (close to 0.1 on a 0 to 1 scale) and there is little systematic relationship with initial relative state governance.

This is consistent with what we learned of combos in interviews during and after the intervention. Combos are principally concerned with the police, who interfere with their drug sales and conduct raids and arrests. As one senior combo member explained to us, "it's important to keep the neighborhood calm. If nothing bad happens, the police don't squeeze us and let us work."

Generally speaking, the gangs seldom impede the activities of civilian officials, community leaders, and regular residents. Indeed, they may have an interest in civilian provision of order if it reduces calls to police. One combo member told us, "If a husband beats his wife, its obligatory that the police will have to come up and intervene. If this happens frequently, the police end up retaking control."[16] Although he was explaining combo involvement in resident affairs, it seems plausible that combos might welcome civilian dispute-resolution services. Harassing or interfering with city officials, meanwhile, could trigger police incursions and the attendant risks of arrest or drug seizure.[17]

6.2 Impacts on legitimacy and reported governance

Table 3 reports the average treatment effects and heterogeneity analysis of our intervention on primary outcomes and absolute levels of legitimacy, for both the state and combo. Overall, the experimental results echo the first-stage results: null effects on average, but positive effects in high-state areas.

[16] Criminal Group Leader 3/29, interview 1/1 [01/05/2021].

[17] We do not want to understate the potential role of combos. First, low initial relative state governance could itself be endogenous to gang strength. Second, residents' heterogeneous, "first-stage" reports of perceptions of and participation in program activities could be mediated by their experience of living under combo governance. Still, combos do not appear to have directly interfered in implementation, a conclusion reinforced by our experimental results on combo governance and legitimacy, reported below.

Table 3: Program impacts on legitimacy and governance: Average treatment effects and heterogeneity by baseline governance quality

			Het. by baseline rel. gov.			
		ATE	Above median	Below median	Diff.	
Dependent variable	Control Mean (SD)	Estimate (CRSE p) [RI p]	Estimate (CRSE p) [RI p]	Estimate (CRSE p) [RI p]	Estimate (CRSE p) [RI p]	N
	(1)	(2)	(3)	(4)	(5)	
Relative state legitimacy index	0.13	0.016	0.050	-0.021	0.071	1,845
	(0.32)	(0.392)	(0.036)	(0.476)	(0.058)	
		[0.294]	[0.101]	[0.326]	[0.112]	
State legitimacy index	0.57	0.013	0.033**	-0.010	0.043**	1,906
	(0.21)	(0.090)	(0.002)	(0.313)	(0.003)	
		[0.146]	[0.029]	[0.272]	[0.031]	
Combo legitimacy index	0.44	-0.002	-0.015	0.012	-0.027	1,845
	(0.28)	(0.877)	(0.394)	(0.613)	(0.357)	
		[0.456]	[0.299]	[0.379]	[0.280]	
Relative state governance index	0.07	-0.025	-0.018	-0.033	0.015	2,314
	(0.31)	(0.068)	(0.378)	(0.063)	(0.581)	
		[0.119]	[0.287]	[0.136]	[0.361]	
State governance index	0.41	-0.012	-0.006	-0.018	0.013	2,362
	(0.26)	(0.162)	(0.630)	(0.109)	(0.434)	
		[0.183]	[0.376]	[0.173]	[0.312]	
Combo governance index	0.35	0.011	0.010	0.012	-0.002	2,316
	(0.28)	(0.389)	(0.598)	(0.451)	(0.926)	
		[0.291]	[0.376]	[0.312]	[0.469]	

Notes: The table reports ITT estimates of program impacts and treatment heterogeneity. Each row is a different dependent variable. We report p-values from cluster robust standard error estimation (CRSE) in parentheses and from randomization inference (RI) in brackets. Both households and businesses were surveyed on governance levels (N=2,379), but only households were surveyed on legitimacy and hence there are fewer observations (N=1,910).

Column 2 reports average treatment effects. We see no signs of significant improvement in legitimacy, and we actually observe a 0.025 *decrease* in relative state governance that is weakly significant (p=0.068) when using CRSEs, though not so when using the more conservative RI p-values (p=0.119).

Turning to the heterogeneity analysis in Columns (3) to (5), relative state legitimacy rose by 0.05 in high-state sectors (p=0.036 using CRSEs and 0.101 using RI). This is equal to 40 percent of the state–combo difference in legitimacy (Column 1) and 9 percent of the average level of absolute state legitimacy (0.57). There is also a small, non-significant decrease in legitimacy in low-state sectors. As a result, the difference between the two subgroups is even larger—0.071, equivalent to 13 percent of the city-wide average. We see the same pattern with absolute state legitimacy, albeit the effects are more robust.[18]

Next we turn to governance, where we see weaker evidence of heterogeneous effects. There is no indication the program increased the frequency with which respondents said the state responded to the 17 forms of disorder—overall or in the high-state subgroup. Perceptions of relative and absolute state governance decline slightly in all sectors, with no statistically significant difference between high- and low-state sectors.[19]

These results are generally robust to alternative estimation approaches, including the omission of control variables, the addition of demographic traits of survey respondents, and

[18] Appendix Table E.1 breaks down legitimacy into its five component questions. The survey asked about the legitimacy of the police and municipal government separately, and we show program impacts on both. In high-state sectors, both mayoral and police legitimacy increase. Interestingly the impact on police is slightly larger and more robust, though the differences between mayoral and police impacts are statistically indistinguishable.

[19] These patterns hold if we break the governance index into its 17 components and into more and less police-related actions, as reported in Appendix Table E.2. We classify the 17 forms of disorder into 8 that are more likely to elicit a call to police or a police officer response, and 9 that are commonly solved by a variety of city and community actors. This is the only way to assess potential differences between police and mayoral staff because, unlike legitimacy, we did not ask all 17 questions for the police and Alcaldía separately (to keep the survey brief). The fall in perceived police responsiveness is slightly greater than the fall in non-police governance, especially in low-state sectors, but none of these differences are statistically significantly different from one another.

The survey also included a number of supplementary measures of efficacy, including the speed of response, ease of accessing services, and the value placed on the actor. We report these in Appendix Table E.3. We see no evidence of that residents perceived an improvement in the speed or ease of contacting the police or mayoral staff, on average or in high-state sectors. There is some evidence that perceived value of the Alcaldía declined in low-state sectors.

the use of a lasso control vector (see Appendix Table E.4). Our heterogeneity analysis is also reasonably robust to alternatives, such as using a predicted measure of absolute baseline state governance.[20] Appendix D.4 also reports formal tests of measurement error, including a survey experiment that uses randomized response techniques to elicit whether sensitive combo-related questions are underreported, and if there is any correlation with treatment status. We find no evidence of bias.

Finally, Table 3 shows no evidence of program impacts on combo governance or legitimacy, even in high-state sectors where the state elicited the largest improvements in state legitimacy and decreases in disorder. We initially hypothesized that better state governance could crowd out combo services, but we see no evidence of this, even in sectors where the intervention was more successful. This is consistent with our qualitative interviews with combo leaders and city officials, discussed above, suggesting that combos did not feel threatened by non-police state presence. Alternatively, it could be that the 20 months intervention was too short or weak to provoke a combo response.

6.3 Impacts on crime and emergency calls

Given that the intervention increased perceived legitimacy (in well-governed sectors) but not perceived state responsiveness, we turn to two sources of objectives security data: reported crimes and security-related calls to the city emergency line. Impacts are again concentrated in initially well-governed sectors.

[20]Using our rich baseline data (including administrative data and leader surveys), we trained a lasso algorithm to predict endline state governance in the 40 control sectors. The resulting prediction is a weighted average of baseline variables that attempts to proxy for absolute rather than relative baseline state governance. Naturally this is a noisier measure than our prespecified heterogeneity term, but if we re-estimate the heterogeneity analysis with an indicator for above/below median predicted state governance, we get qualitatively similar results (see Appendix Table E.5). The results suggest a statistically significant increase in perceived state governance in high-state sectors. And the legitimacy results are similar though marginally less significant (p=0.106 with RI and p=0.049 with CRSEs.)

In addition to our prespecified high- and low-state comparisons, we also committed to report treatment effects in the four major quartiles, as seen in Appendix Table E.6. With just 20 sectors per subgroup, however, that analysis is under-powered. Nonetheless, the legitimacy results persist, with some evidence that the backlash is concentrated in the lowest quartile. The patterns we observe in crime and calls to police are generally consistent with the high- and low-state analysis described below. The governance results are

Table 4: Program impacts on crime index components: Average treatment effects and heterogeneity by baseline governance quality

Dependent variable	Control Mean (SD)	ATE Estimate (CRSE p) [RI p]	Het. by baseline rel. gov.: Above median Estimate (CRSE p) [RI p]	Het. by baseline rel. gov.: Below median Estimate (CRSE p) [RI p]	Het. by baseline rel. gov.: Difference Estimate (CRSE p) [RI p]	N
	(1)	(2)	(3)	(4)	(5)	
Sentence-weighted crime index	0.35	-0.061*	-0.137**	0.023	-0.160**	80
	(0.26)	(0.027)	(0.001)	(0.424)	(0.002)	
		[0.071]	[0.027]	[0.289]	[0.028]	
Homicides	0.04	0.029**	0.025	0.032*	-0.007	80
	(0.05)	(0.001)	(0.043)	(0.013)	(0.710)	
		[0.024]	[0.112]	[0.061]	[0.407]	
Vechicle thefts	0.33	-0.040	-0.123*	0.053	-0.176*	80
	(0.30)	(0.317)	(0.031)	(0.264)	(0.017)	
		[0.243]	[0.091]	[0.218]	[0.060]	
Thefts and robbery	1.44	-0.463**	-0.886**	0.009	-0.895**	80
	(1.46)	(0.007)	(0.001)	(0.953)	(0.003)	
		[0.030]	[0.021]	[0.480]	[0.035]	
Assaults	0.64	-0.104**	-0.146**	-0.057	-0.089	80
	(0.38)	(0.004)	(0.007)	(0.249)	(0.238)	
		[0.035]	[0.047]	[0.240]	[0.231]	

Notes: The table reports summary statistics and treatment effects for the sentence-weighted crime index in Table 3 and its four main components. The index is standardized to have zero mean and unit standard deviation. We report p-values from cluster robust standard error estimation (CRSE) in parentheses and from randomization inference (RI) in brackets.

For reported crime, our index falls by 0.137 in high-state sectors—a 40% decline, significant at the 5 percent level. The divergence between above- and low-state sectors is even greater, a 0.16 decline. Proportionally, these are large declines for most crime types—vehicle thefts, other thefts and robber, and assault. Curiously, however, we see a rise in homicides overall in treated areas. We must treat all index component analyses as suggestive, however, and we have not adjusted standard errors for multiple hypothesis tests.

Note that these reductions in crime are likely to arise from differential reporting of crime in treated communities. Residents in initially well-governed treated sectors view the state as more legitimate, and so if anything should be more willing to report crimes to the state. Moreover, the intervention explicitly educated communities on the police code and facilitated semi-annual meeting between the community and local police commander, thus making them more familiar with reporting requirements. In principle, these factors could increase crime reporting rates in treated sectors, leading us to understate treatment effects.

The evidence from security-related calls further suggests that, in high-state sectors, municipal staff or the community itself is either dealing with everyday street disorder without the police, or successfully prevented forms of disorder. In high-state treated sectors, calls fall by 63 relative to a control mean of 136—a 55 percent decline, significant at the 5 percent level. There is no evidence of improvement in the low-state sectors. We see this differential across every category of call, except for the very small number of firearm-related altercations (See Column 5). The largest decline (and the only statistically significant component) is in calls regarding unarmed street fights and domestic abuse.

6.4 Impacts on a summary index

To avoid concerns of multiple hypothesis testing and non-prespecified outcomes, we construct a family index of all four measures—relative legitimacy, relative governance, security-related calls, and reported crime—and test for average and heterogeneous treatment effects, in Ap-

less consistent, and adverse effects are not necessarily concentrated in the lowest quartiles.

Table 5: Impacts of treatment on security-related emergency calls

Dependent variable	Control Mean (SD)	ATE Estimate (CRSE p) [RI p]	Het. by baseline rel. gov.: Above median Estimate (CRSE p) [RI p]	Below median Estimate (CRSE p) [RI p]	Difference Estimate (CRSE p) [RI p]	N
	(1)	(2)	(3)	(4)	(5)	
Security-related emergency calls	135.76	-34.969**	-63.250***	-3.519	-59.731**	80
	(102.30)	(0.008)	(0.000)	(0.803)	(0.003)	
		[0.027]	[0.009]	[0.435]	[0.048]	
Physical altercations	93.34	-18.414**	-35.583***	0.678	-36.261**	80
	(58.36)	(0.002)	(0.000)	(0.942)	(0.005)	
		[0.043]	[0.004]	[0.481]	[0.044]	
Narcotics related incidents	30.97	-15.870*	-24.909*	-5.819	-19.090	80
	(60.11)	(0.123)	(0.078)	(0.409)	(0.064)	
		[0.059]	[0.086]	[0.269]	[0.249]	
Armed incidents	11.45	-0.684	-2.758*	1.622	-4.381*	80
	(7.97)	(0.515)	(0.033)	(0.341)	(0.042)	
		[0.337]	[0.100]	[0.262]	[0.083]	
Knife related incidents	9.25	-0.968	-2.772**	1.038	-3.810*	80
	(6.95)	(0.199)	(0.003)	(0.381)	(0.014)	
		[0.214]	[0.035]	[0.287]	[0.054]	
Firearm related incidents	2.20	0.284	0.014	0.584	-0.570	80
	(2.44)	(0.585)	(0.982)	(0.494)	(0.580)	
		[0.352]	[0.493]	[0.343]	[0.357]	

Notes: This table reports the total number of resident calls to the police emergency line over 20 months, including all calls made within each sector plus a 125 meter buffer zone around the sector. Calls are only geolocated within the city if the police actually respond to the call, meaning we cannot track impacts on unmet calls. But administrative records suggest that more than 97% of calls receive a response, and so are unlikely to affect out results. We report p-values from cluster robust standard error estimation (CRSE) in parentheses and from randomization inference (RI) in brackets.

pendix Table E.4. The results are largely consistent with the patterns discussed above: no evidence of an average treatment effect, but robust evidence of improvements in the high-state areas.

7 Discussion and conclusions

Cities around the world have patchworks of high state control and legitimacy abutting areas where the state is weak and mistrusted. In Latin America, the issue takes on special importance, as tens if not hundreds of millions also live under some form of criminal governance in addition to that of the state (Uribe et al. 2022).

Why these disparities persist is a puzzle, as is what to do about it. The most common response has been to expand the coercive power of the state—professionalize and expand their police in some cases, while pursuing militarized and heavy-handed *mano dura* approaches in others. This paper examines one city's attempt to tackle the problem noncoercively—through better community-government communication, enhanced municipal services, and increased engagement with civilian street-level bureaucrats. It contributes rare experimental evidence on a key question for statebuilding: whether enhanced everyday governance (and other alternatives to coercive forces) can meaningfully improve order and state legitimacy.

Our experimental results suggest that intensifying non-police state presence improved legitimacy and security, but only where the state was already present and providing relatively good governance. In many settings, the returns to investment are thought to be diminishing: each additional unit of labor, capital, or technology applied to production (a manufacturing plant, a public service, a political campaign) has positive but less than proportional returns. This was our initial expectation—that a state's first efforts to bring more visibility and services to a neighborhood would bring the most attention, impact, and gratitude. But there are also settings with increasing returns—early investments lead to little (or even negative) changes in output, until some threshold is reached.

One possibility is that, when it comes to statebuilding, this threshold comes from state capacity. In Medellín's *Operación Convivencia*, treatment compliance—the actual delivery of enhanced attention and governance services—appears to have varied with the degree to which the police and Alcaldía were already present and addressing problems in the neighborhood. Despite a costly and good-faith effort to apply a uniform "dose" of enhanced governance, both the liaisons and the broader state apparatus seem to have fallen short in communities where their initial presence was relatively low. If the state made promises it then tried but failed to deliver on in especially poorly-served neighborhoods, it is not surprising that residents' perceptions of its legitimacy plateaued or fell.

The sources of increasing returns in firms and economic development are well-documented. Increasing returns in politics are implicit in claims of path dependence (Pierson 2000), but they are less well theorized and documented. Understanding whether they exist in state-building and why is an important area for future research. In this instance, we hypothesize three potential sources of increasing returns to public investment and penetration.

First, there could be start-up costs or entry barriers to governing which, if not overcome, make project execution all but impossible. For example, we were concerned that non-state actors could thwart or undermine weak state efforts to enter their territory. This would be an example of entry barriers, although not one we encountered in Medellín. Other start-up costs are more plausible in this scenario. There could be a threshold for general awareness—the state must undertake some minimum level of effort before the average resident can be expected to notice the activity and change their beliefs about the state. Alternatively, levels of state effort below some threshold could draw attention to inequities and failures, and harm the state's reputation. There are cases of this in the literature. For instance, Gottlieb (2016) argues that a civics education program in Mali raised citizen expectations of politicians and led to greater survey-based willingness to sanction leaders. In Liberia, Blair et al. (2019) and Karim (2020) find that policing interventions can lower state and police legitimacy when they raised expectations beyond the capacity to deliver.

A second possibility is that low initial relative state governance reflects neighborhood characteristics—such as low social cohesion or marginalized ethnic or social background—that reduce communities' ability to mobilize and hold officials accountable (Moncada 2024). In this scenario, implementation efforts fail not because bureaucrats and agents are unable to deliver, but because they lack the incentives. When state agents are tasked with delivering services to more marginalized, disorganized, or low social capital areas, those communities fall to the back of the priority list or are outright neglected. This logic underlies Vargas (2016) account of gun violence in Chicago—there are many poor and marginalized neighborhoods, and the ones that receive violence prevention services and enjoy calm are those with greater social capital and political connections.

A third possibility is that lower initial relative state governance is related to other unobserved confounders that affect their interactions with and perceptions of statebuilding efforts. For example, residents in such sectors might have lower demand for state services, leading to lower delivery. We found no evidence for this dynamic in Medellín. Rather, local leaders, residents and businesspersons consistently demanded for more police and mayoral attention in our qualitative interviews.

These mechanisms are not mutually exclusive. They are complementary. Marginalization, community self-reliance, and hostility toward the state could endogenously generate start-up costs and entry barriers.

Granted, these are big conjectures—well beyond what a single experiment of 80 neighborhoods in one city can demonstrate. Its generalizability remains to be tested (though there is no shortage of settings in which to do so). Still, experimental evaluation of statebuilding techniques has to begin somewhere. The first community-level randomized evaluation of a civilian security intervention will have natural limitations. If nothing else, Operación Convivencia illustrates the viability of community-level experimentation and rigorous evaluation of alternative statebuilding strategies, and its potential contribution to both theory and practice. Just as importantly, its results forced us to reject our working hypotheses and question the theoretical assumptions behind them. This is an essential part of the scientific method, generating new theories and hypotheses for future research.

These new hypotheses could help explain one of the most salient features of city life: the gaping inequality in governance in neighborhoods often no more than a few blocks from one another. Increasing returns could produce perverse incentives, ones that lead bureaucrats and elected officials to focus their effort and scarce resources in the areas where visible, short-term progress is easiest to achieve, neglecting neighborhoods that are already behind. As the world becomes increasingly urban, progress towards understanding this "neglect trap" is arguably one of the most important frontiers of statebuilding research.

References

Acemoglu, Daron, Ali Cheema, Asim I Khwaja, and James A Robinson. 2020. "Trust in state and nonstate actors: Evidence from dispute resolution in Pakistan". *Journal of Political Economy* 128 (8): 3090–3147.

Akerlof, George A. and Janet L Yellen. 1994. "Gang Behavior, Law Enforcement, and Community Values". In H. J. Aaron, T. E. Mann, and T. Taylor (Eds.), *Values and Public Policy*, pp. 173–209. Washington, DC: The Brookings Institution.

Arias, Enrique Desmond. 2017. *Criminal Enterprises and Governance in Latin America and the Caribbean.* New York: Cambridge University Press.

Arjona, Ana. 2016. *Rebelocracy.* Cambridge University Press.

Bai, Yuehao. 2022. "Optimality of matched-pair designs in randomized controlled trials". *American Economic Review* 112 (12): 3911–40.

Beath, Andrew, Fotini Christia, and Ruben Enikolopov. 2020. "Winning Hearts and Minds Through Development? Evidence From A Field Experiment in Afghanistan". *Working Paper*.

Berman, Eli, Joseph H Felter, Jacob N Shapiro, and Erin Troland. 2013. "Modest, Secure, and Informed: Successful Development in Conflict Zones". *American Economic Review* 103 (3): 512–17.

Berman, Eli and David D Laitin. 2008. "Religion, Terrorism and Public Goods: Testing the Club Model". *Journal of Public Economics* 92 (10-11): 1942–1967.

Berman, Eli, Jacob N Shapiro, and Joseph H Felter. 2011. "Can Hearts and Minds be Bought? The Economics of Counterinsurgency in Iraq". *Journal of Political Economy* 119 (4): 766–819.

Blair, Robert and Michael Weintraub. 2021. "Military policing exacerbates crime and may increase human rights abuses: A randomized controlled trial in cali, colombia". *Working paper*.

Blair, Robert A and Pablo Kalmanovitz. 2016. "On the Rights of Warlords: Legitimate Authority and Basic Protection in War-Torn Societies". *American Political Science Review* 110 (3): 428–440.

Blair, Robert A., Sabrina M. Karim, and Benjamin S. Morse. 2019. "Establishing the rule of law in weak and war-torn states: Evidence from a field experiment with the liberian national police". *American Political Science Review* 113 (3): 641–657.

Blattman, Christopher, Gustavo Duncan, Benjamin Lessing, and Santiago Tobón. 2024. "Gang Rule: Understanding and Countering Criminal Governance". *Review of Economic Studies (forthcoming)*.

Blattman, Christopher, Donald Green, Daniel Ortega, and Santiago Tobón. 2021. "Place-Based Interventions at Scale: The Direct and Spillover Effects of Policing and City Services on Crime". *Journal of the European Economic Association* 19 (4): 2022–2051.

Blattman, Christopher, Alexandra C. Hartman, and Robert A Blair. 2014. "How to Promote Order and Property Rights Under Weak Rule of Law? An Experiment in Changing Dispute Resolution Behavior Through Community Education". *American Political Science Review*: 100–120.

Braga, Anthony A., Brandon C. Welsh, and Cory Schnell. 2015. "Can policing disorder reduce crime? A systematic review and meta-analysis". *Journal of Research in Crime and Delinquency* 52 (4): 567–588.

Bruhn, Miriam and David McKenzie. 2009. "In pursuit of balance: Randomization in practice in development field experiments". *American Economic Journal: Applied Economics* 1 (4): 200–232.

Cammett, Melani and Lauren M MacLean. 2014. *The Politics of Non-State Social Welfare*. Cornell University Press.

Carter, Danielle. 2013. *Non-state security, state legitimacy and political participation in South Africa*. Michigan State University.

Cassidy, Tali, Gabrielle Inglis, Charles Wiysonge, and Richard Matzopoulos. 2014. "A systematic review of the effects of poverty deconcentration and urban upgrading on youth violence". *Health and Place* 26 : 78–87.

Chalfin, Aaron and Justin McCrary. 2017. "Criminal deterrence: A review of the literature". *Journal of Economic Literature* 55 (1): 5–48.

Collazos, Daniela, Eduardo García, Daniel Mejía, Daniel Ortega, and Santiago Tobón. 2021. "Hot Spots Policing in a High-Crime Environment: An Experimental Evaluation in Medellin". *Journal of Experimental Criminology* 112 : 473—-506.

Collier, David, Fernando Daniel Hidalgo, and Andra Olivia Maciuceanu. 2006, oct. "Essentially Contested Concepts: Debates and Applications". *Journal of Political Ideologies* 11 (3): 211–246.

Cruz, José M. 2011. "Government Responses and the Dark Side of Gang Suppression in Central

America". In T. Bruneau, L. Dammert, and E. Skinner (Eds.), *Maras: Gang Violence and Security in Central America*, pp. 137–158. Austin: University of Texas Press.

Cruz, José M. and Angelica Durán-Martínez. 2016. "Hiding violence to deal with the state: Criminal pacts in El Salvador and Medellin". *Journal of Peace Research* 53 (2): 197–210.

Duflo, Esther and Abhijit Banerjee. 2011. *Poor economics.* PublicAffairs.

Gambetta, Diego. 1996. *The Sicilian Mafia: the Business of Private Protection.* Harvard University Press.

Gerber, Alan S and Donald P Green. 2012. *Field experiments: Design, analysis, and interpretation.* W.W. Norton.

Gilley, Bruce. 2009. *The right to rule: How states win and lose legitimacy.* Columbia University Press.

Gonzalez, Robert and Sarah Komisarow. 2020. "Community monitoring and crime: Evidence from chicago's safe passage program". *Journal of Public Economics* 191 : 104250.

Gottlieb, Jessica. 2016. "Greater expectations: A field experiment to improve accountability in mali". *American Journal of Political Science* 60 (1): 143–157.

Gurr, Ted R. 1971. "Why men rebel".

Hartman, Alexandra C, Robert A Blair, and Christopher Blattman. 2021. "Engineering informal institutions: Long-run impacts of alternative dispute resolution on violence and property rights in liberia". *The Journal of Politics* 83 (1): 381–389.

Henn, Soeren J. 2021. "Complements or Substitutes? How Institutional Arrangements Bind Chiefs and the State in Africa".

Herbst, Jeffrey. 2006. "Population Change, Urbanization, and Political Consolidation". In R. Goodin and C. Tilly (Eds.), *The Oxford Handbook of Contextual Political Analysis*, pp. 649–663. New York: Oxford University Press.

Herbst, Jeffrey. 2014. *States and Power in Africa: Comparative Lessons in Authority and Control.* Princeton: Princeton University Press.

Horowitz, Donald L. 2000. *Ethnic groups in conflict, updated edition with a new preface.* Univ of California Press.

Kalyvas, Stathis N. 2006. *The Logic of Violence in Civil War.* Cambridge University Press.

Karim, Sabrina. 2020. “Relational state building in areas of limited statehood: Experimental evidence on the attitudes of the police”. *American political science review* 114 (2): 536–551.

Krasner, Stephen D and Thomas Risse. 2014. “External actors, state-building, and service provision in areas of limited statehood”. *Domestic Politics and Norm Diffusion in International Relations*: 197.

Lake, Milli. 2022. “Policing insecurity”. *American Political Science Review* 116 (3): 858–874.

Lessing, Benjamin. 2020. “Conceptualizing Criminal Governance”. *Perspectives on Politics*: 1–20.

Lessing, Benjamin, Douglas. Block, and Elayne Stecher. 2019. “Criminal Governance in Latin America: an Empirical Approximation”. *Working Paper*.

Levi, Margaret, Audrey Sacks, and Tom Tyler. 2009. “Conceptualizing legitimacy, measuring legitimating beliefs”. *American behavioral scientist* 53 (3): 354–375.

Leyva, Santiago. 2010. “El proceso de construcción de estatalidad local (1998-2009): ¿La clave para entender el cambio en Medellín?”. In M. Hermelín, A. Echeverri, and J. Giraldo (Eds.), *Medellín: Medio Ambiente, Urbanismo, Sociedad*, pp. 271–293. Medellín: EAFIT.

MacKinnon, James G, Morten Ørregaard Nielsen, and Matthew D Webb. 2023. “Cluster-robust inference: A guide to empirical practice”. *Journal of Econometrics* 232 (2): 272–299.

Magaloni, Beatriz, Edgar Franco-Vivanco, and Vanessa Melo. 2020. “Killing in the Slums: Social Order, Criminal Governance, and Police Violence in Rio de Janeiro”. *American Political Science Review* 114 (2): 552–572.

Mampilly, Zachariah Cherian. 2012. *Rebel rulers: Insurgent governance and civilian life during war.* Cornell University Press.

Melnikov, Nikita, Carlos Schmidt-Padilla, and Maria Micaela Sviatschi. 2020. “Gangs, Labor Mobility, and Development: The Role of Extortion in El Salvador”.

Mnookin, Robert H. 1998. *Alternative dispute resolution.* Harvard Law School.

Moncada, Eduardo. 2024. “Criminal Competition and Collective Political Mobilization”.

Moncada, Juan José, Carolina Lopera, Natalia Maya, Claudia Cadavid, and Lina Zuluaga. 2018. *La Extorsión en Medellín Como Fenómeno del Orden Social, Poder Político y Control Territorial.* Medellín: Alcaldía de Medellín.

Morris, Kevin T and Kelsey Shoub. 2024. “Contested killings: the mobilizing effects of community contact with police violence". *American political science review* 118 (1): 458–474.

Owens, Emily. 2019. "Economic approach to de-policing". *Criminology & Public Policy* 18 : 77.

Pierson, Paul. 2000. "Increasing returns, path dependence, and the study of politics". *American Political Science Review* 94 (2): 251–267.

Risse, T. 2011. "Policies and politics in areas of limited statehood. introduction and overview". *Governance without State*: 1–35.

Risse, Thomas and Eric Stollenwerk. 2018. "Legitimacy in areas of limited statehood". *Annual Review of Political Science* 21 : 403–418.

Skaperdas, Stergios. 2001. "The Political Economy of Organized Crime: Providing Protection When the State Does Not". *Economics of Governance* 2 (3): 173–202.

Skaperdas, Stergios and Constantinos Syropoulos. 1996. *Gangs as Primitive States*, pp. 61–81. Cambridge University Press.

Skarbek, David. 2011. "Governance and Prison Gangs". *American Political Science Review* 105 (4): 702–716.

Staniland, Paul. 2012. "States, insurgents, and wartime political orders". *Perspectives on politics* 10 (2): 243–264.

Tilly, Charles. 1990. *Coercion, Capital, and European States, AD 990-1990.* Blackwell.

Tyler, Tom R. 2003. "Procedural Justice, Legitimacy, and the Effective Rule of Law". *Crime & Justice* 30 : 283–358.

Tyler, Tom R. 2004, may. "Enhancing Police Legitimacy". *The Annals of the American Academy of Political and Social Science* 593 (1): 84–99.

Uribe, Andres, Benjamin Lessing, Douglas Block, Elayne Stecher, and Noah Schouela. 2022. "Criminal Governance in Latin America: an Initial Assessment of its Extent and Correlates".

Van der Windt, Peter, Macartan Humphreys, Lily Medina, Jeffrey F Timmons, and Maarten Voors. 2019. "Citizen Attitudes Toward Traditional and State Authorities: Substitutes or Complements?". *Comparative Political Studies* 52 (12): 1810–1840.

Vargas, Robert. 2016. *Wounded city: Violent turf wars in a Chicago barrio.* Oxford University Press.

Weber, Max. 1946. *Politics as a Vocation*, pp. 77–128. Oxford University Press.

Wedeen, Lisa. 2015. *Ambiguities of Domination* (2nd ed.). Chicago: The University of Chicago Press.

Weil, David N. 2008. *Economic Growth* (2nd ed.). Boston: Addison-Wesley.

Appendix

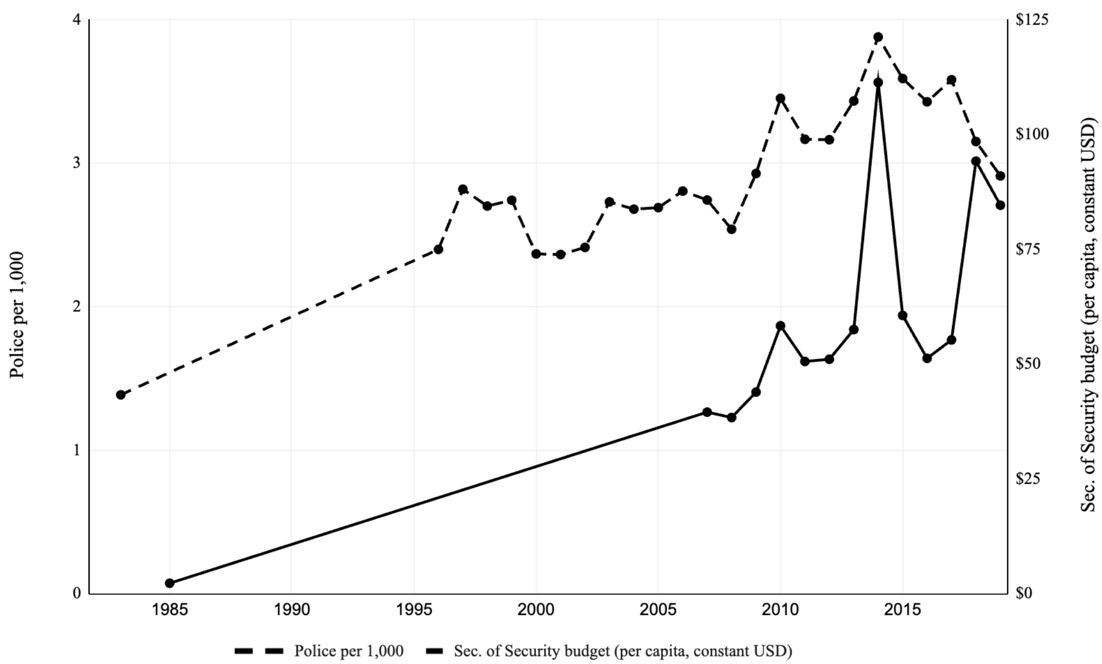
A Spending on state security in Medellín over time

While the intervention we study is an intensive one, it is not an impossible margin on which to increase a security and services budget over the course of time.

The civilian Secretariat of Security emerged and grew over the last three decades. Our archival work uncovered a budget of $2.23 per capita for 1985, in constant 2019 USD. By 2007 (the next available year we could find data) spending was $39.50 per capita. A decade later the figure was above $50 (Figure A.1).

We have seen similarly large national investments in policing. For instance, there were 1.4 police officers per 1,000 residents in 1983. The national government doubled to 2.8 by 1997, the next year for which there is archival data (**?**). We have annual data thereafter, however, and by 2017 the city had 3.6 police officers per 1,000 people (Figure A.1).

Figure A.1: Growth in state security and spending personnel



Notes: The figure depicts the evolution of police per 1,000 residents (reported on the left vertical axis) and the evolution of per capita city expenditures on security and protection (reported on the right vertical axis). Data are from the City of Medellín.

B Research ethics and transparency

B.1 Human subjects protections of criminal group members

Our qualitative interviews with criminal group members are naturally sensitive, especially those conducted in prison.

We believe criminal group members spoke to us for several reasons: pride; respite from boredom; interest in speaking with professors; the fact that they were already prosecuted; the fact that we were not asking about indictable information; and a hope that this would further their efforts for a peace process with the government. Generally speaking, we spoke with middle- and high-ranking leaders who were extremely powerful and fully aware of the benefits and risks of speaking to us.

Our sources may also have had their own agendas. Some remarked that the government underestimated their strength, that this interfered with bargaining, and that we could resolve this as we seemed to have a more accurate understanding of the situation. Others openly hoped for some kind of negotiation with the government, and may have seen our project as consistent with those objectives.

Our subjects also seemed to be drawn from the members most interested in a chance to exhibit their expertise and insights. Occasionally, we were referred to people who were highly suspicious and unwilling to speak. The ones who did talk to us seemed flattered by academic attention, and several indicated an interest in being the subject of research. In prison, our interviews also offer subjects a respite from routine.

We had several strategies for maintaining full voluntary consent as well as the confidentiality of criminal group members. Above all, with all interviewees, we were transparent about our research aims and work with the government. Interviews in prison were fully voluntary, and prison authorities appeared indifferent to the cooperation or participation of potential interviewees.

We also made every effort to preserve anonymity and confidentiality, while advising subjects in consent scripts of the potential limits to our ability to do so. Finally, we consulted extensively with the human subjects committees of our institutions, and we obtained written support and assurances of noninterference from several authorities, including the mayor and ministry of justice.

In general, we think the information we collected is relatively reliable. One reason is that gang organization and rule are not sensitive subjects or prosecutable offenses. In addition, we sought to validate our observations with multiple sources. For most topics we discuss we have at least 2–3 sources between gang members, ex-members, and experts. Nonetheless, for obvious reasons, we cannot trust these qualitative data entirely.

B.2 Experimental intervention

For the experimental intervention, we worked with the government of Medellín to design a program that simply intensified a set of "everyday" governance services that the city was already providing on a wide basis. What's more, this intensified model was based on an existing models (described in the main text). Thus, the experiment merely formalized the evaluation of well-tested municipal services and practices.

In order to minimize the risk that the intervention would reduce the city services received by some neighborhoods, the city hired new personnel as liaisons and to supplement the city's central task force.

Of course, there may still have been displacement or crowding out of services from other neighborhoods. Is this an ethical concern? Arguably, this kind of variation and redistribution of services is a normal feature of urban policy decision-making and improvement, much like any place-based criminal justice or municipal policy approach. As elected officials they were seeking to learn how to improve order and legitimacy in underserved neighborhoods. Typically policymakers do so in an informal and iterative way. A more formal experiment helped them do so in more structured and informed way. The same officials make tradeoffs about where to allocate scarce budgets every day, and most of these decisions (which gangs to target, or where to improve infrastructure) typically create winners and losers. Thus, even if the experiment generated major crowding out effects, this is well within the normal bounds of democratic urban decision-making. As it happens, the effort to minimize spillovers and avoid disruption to non-treatment blocks meant this intervention likely had fewer trade-offs and crowding out effects than the usual policy decisions.

C Conceptual framework

C.1 Cournot competition in local governance

To understand why the intervention could crowd combos out of local governance, we can look at the intervention through the lens of imperfect competition for governance services. Any model of imperfect competition should produce similar comparative statics, but we illustrate with Cournot competition, where each side chooses a fixed quantity of protection services to provide and let prices clear the market.[21]

Setup In each neighborhood, a state s and a gang g compete to sell protection in quantities q_g and q_s. Each organization i chooses q_i to maximize their respective pay-off, and each has constant marginal cost c_i. (Here i can either be the state or the gang, and in what follows, j represents a general form of notation for the competing organization.) Products are differentiated, and the price of each one is given by the linear inverse demand function $p_i = a_i - \beta q_i - \gamma q_j$. Here, $\gamma \in (0, 1]$ since the services offered by both organizations are substitutes, and $\beta > 0$ for downward-sloping demand. The pay-off for each organization is $V_i = p_i q_i - c_i q_i$. For simplicity, we assume an interior solution.

[21] Note that Cournot fits some of our stylized facts well—especially that governing requires investments and advanced commitments, and that it is hard to adjust output capacity quickly.

Nash Equilibria The best response function for each organization are derived as follows:

$$\begin{aligned}
\max_{q_i} V_i &= (a_i - \beta q_i - \gamma q_j)q_i - c_i q_i \\
\frac{\partial V_i}{\partial q_i} &= a_i - 2\beta q_i - \gamma q_j - c_i = 0 \\
q_i^* &= \frac{a_i - c_i}{2\beta} - \frac{\gamma}{2\beta} q_j
\end{aligned}$$

Replacing values we obtain (for each organization):

$$q_i^* = \frac{2\beta(a_i - c_i) - \gamma(a_j - c_j)}{(4\beta^2 - \gamma^2)}$$

Comparative statics We are interested in how the quantity of services supplied by the gang behave in response to any increase in state governance: $\frac{\partial q_i^*}{\partial q_j}$. To obtain this comparative static, we begin by defining:

$$G(q_i, q_j) \equiv \frac{\partial V_i}{\partial q_i} = a_i - 2\beta q_i - \gamma q_j - c_i$$

which is a continuously differentiable function from $\mathbb{R}^2 \to \mathbb{R}$. At the optimum, we know:

$$G(q_i^*, q_j^*) = a_i - 2\beta_i q_i^* - \gamma q_j^* - c_i = 0.$$

Since $-2\beta \neq 0$, we can use the implicit function theorem to obtain our main comparative static:

$$\frac{\partial q_i^*}{\partial q_j} = -\frac{\partial G(q_i, q_j)/\partial q_j}{\partial G(q_i, q_j)/\partial q_i} = -\frac{\gamma}{2\beta}$$

Since the two services are not complements, this comparative static implies that increases in one duopolist's supply of protection will reduce the other's.

C.1.1 Cournot competition with benefits to governing

In the simple model above, increases in the quantity supplied by the state would mainly come from reductions in the state's marginal cost of providing these goods. One way to conceive the experimental intervention is an exogenous investment by the state in lowering the marginal cost of providing governance services. Another way to view the intervention, however, is the result of an exogenous increase in the value the state places on being the market leader in that neighborhood, or even a monopolist. To illustrate this, we introduce a new term to the utility function.

Setup As above, but now the payoff for each organization is $V_i = p_i q_i - c_i q_i + \rho(q_i, q_j)\pi_i$, where $\rho(q_i, q_j)\pi_i$ represents each player's returns to loyalty, legitimacy, and control of the neighborhood.

Set up this way, π_i is the return to full control of the block. For example, π_s includes electoral rewards, achievement of policy aims, or preferences for dominance and citizen loyalty.

Meanwhile, $\rho(\cdot)$ scales each organization's ability to capture, retain, or enjoy these benefits. We can think of it as the share of π_i each player enjoys, one that is increasing in own governance and decreasing in the other's, such that: $\frac{\partial \rho(q_i,q_j)}{\partial q_i} > 0 > \frac{\partial \rho(q_i,q_j)}{\partial q_j}$. Importantly, however, we remain agnostic here about whether $\rho(\cdot)$ exhibits increasing or decreasing returns to own and other's governance provision.

Nash Equilibria For simplicity, we assume an interior solution. We can derive the best response function for each organization:

$$
\begin{aligned}
\max_{q_i} V_i &= (a_i - \beta q_i - \gamma q_j)q_i - c_i q_i + \rho(q_i, q_j)\pi_i \\
\frac{\partial V_i}{\partial q_i} &= a_i - 2\beta q_i - \gamma q_j - c_i + \frac{\partial \rho(q_i, q_j)}{\partial q_i}\pi_i = 0 \\
q_i^* &= \frac{a_i - c_i + \frac{\partial \rho(q_i,q_j)}{\partial q_i}\pi_i}{2\beta} - \frac{\gamma}{2\beta}q_j
\end{aligned}
$$

We obtain an identical best response function for the other organization analogously, and replacing values we obtain:

$$
q_i^* = \frac{2\beta(a_i - c_i) - \gamma(a_j - c_j) + \left(2\beta\frac{\partial \rho(q_i,q_j)}{\partial q_i}\pi_i - \gamma\frac{\partial \rho(q_i,q_j)}{\partial q_j}\pi_j\right)}{(4\beta^2 - \gamma^2)}
$$

with an identical function for q_j^*.

The state's equilibrium level of governance services supplied is increasing in the value they place on neighborhood control, π_i, and their expected returns to investment in citizen loyalty and neighborhood control, $\frac{\partial \rho(q_i,q_j)}{\partial q_i}$.

D Experimental procedures

D.1 Preanalysis plan and deviations

We preregistered our design, outcomes, estimation, and heterogeneity analysis in the social science registry in April 2018, as the intervention was launched. We refined and re-registered the design in October 2019, prior to final data collection, as a JDE registered report. An anonymized copy of this pre-analysis plan is enclosed.

There were only two deviations from this plan. First, we changed how we conceptualize the intervention. As described in in Section 4 and Appendix C, we initially believed that state governance could not only raise state performance and legitimacy, but also that this could crowd out combo governance. We still explore this possibility. But as the intervention unfolded, and we had a chance to interact with the liaisons and task force over two years, our understanding of their work evolved. Liaisons generally did not try to crowd out the gang, and gangs generally did not react to a non-police intervention. Instead, we recognized that *convivencia* was more focused on fostering local order, and that our analysis should try to assess this by considering crime and calls to the police as secondary analyses. What's more, following the launch of the intervention, policing debates in Colombia and the United States highlighted the importance of this intervention as a civilian-led approach to neighborhood order. Our current framing—a focus on the returns to urban statebuilding, especially noncoercive efforts—tries to capture all of these questions at once. Throughout we have tried to balance a transparent account of ex ante design and hypotheses with a coherent and forward-looking set of ex post lessons learned and theoretical implication.

Second, we initially committed to report treatment effects for the four major quartiles of baseline relative governance. With just 10 treated and 10 control sectors per subgroup, however, that analysis is under-powered and noisy. Thus we concentrate on one of these quartiles—the median—reporting treatment effects in sectors above and below this median level of initial relative governance. Appendix Table E.6 reports the other quartiles for transparency.

D.2 Sample, matching, randomization, and balance

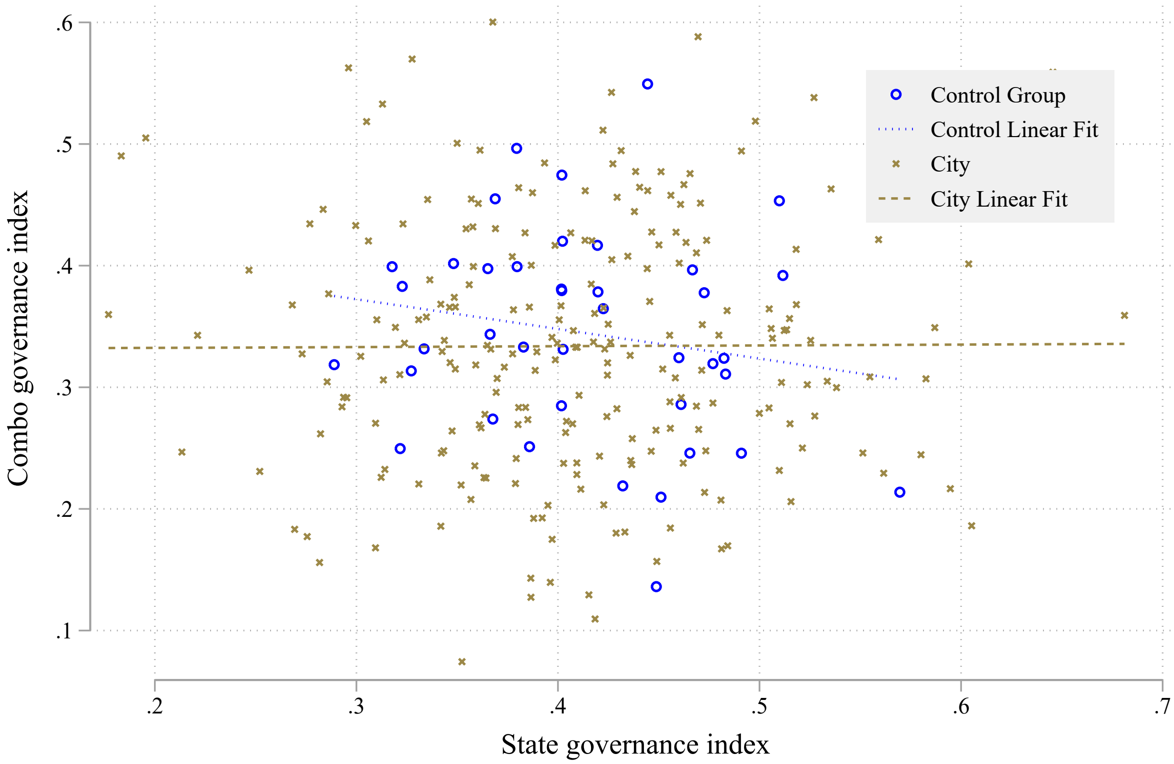
The experimental sectors were selected purposefully by the Secretariat of Security, but they are broadly representative of the city's neighborhoods in terms of their demographics, geographic features, and variation of state and combo governance. Figure D.1 plots experimental sectors against barrio averages.

Note that the *experimental sample* does not overlap with the city sample. There are less than 400 blocks in the experimental sample—less than 2.5 percent of all city blocks. Could the treatment influence outcomes in the city sample? We test for evidence of such spillovers in Appendix Table D.3, below, and find no evidence of them.

For randomization, we grouped the 80 sectors into 40 matched pairs using the four baseline measures of security and governance. We formed pairs so as to minimize the Mahalanobis distance between the values of the four covariates, and then randomly assigned one unit in each pair to treatment. Such matched pair designs can maximize statistical precision, especially in a relatively small sample, although it requires the assumption of common support (Bruhn and McKenzie 2009; Bai 2022).

This approach works best when the baseline measures are prognostic of potential outcomes. For one of the four measures, we constructed an *Index of reported crime* from the previous decade, weighted by severity using criminal sentences. Otherwise, however, the sectors had relatively little baseline data. They are small informal neighborhoods, much smaller than the barrio, and there is no administrative reporting at the sector level. Thus, in order to assess baseline security governance in the sectors, we used the baseline survey of leaders, described in Section 5. In addition to matching on the *Relative state governance* measure, we also used survey data on *Relative state-combo visibility* on the street, and *Perceptions of local security and drug use.*

Figure D.1: Comparison of the experimental and city (representative) sample of blocks in 2019: State and combo governance levels



Notes: The figure plots average 2019 state and combo governance levels in each city barrio as well as the 40 experimental control sectors. We omit treated sectors because the 2019 survey is post-treatment. The dashed lines are lines of best fit for the two samples. The experimental sectors are widely distributed, much like the city barrios, though there are slightly more high combo/low state governance areas in the experimental sample.

This approach produced the expected degree of balance along baseline covariates, as seen in Table D.1. Given the small sample size, we report p-values from randomization inference. In addition to showing balance on the matching variables, the table also reports a selection of additional baseline variables that we did not use in the matching algorithm. Only one variable—the baseline crime index—shows a chance imbalance. Similarly, Table D.2 shows the expected degree of balance for our prespecified subgroups.

Table D.1: Baseline summary statistics and test of balance

Covariate	Means		Regression Difference		N
	Control	Treated	Coeff	RI p-value	
Baseline indices used for matching (standardized)					
Standardized index of frequency of combo visibility	0.02	-0.03	-0.05	0.34	2379
Standardized values of relative state-combo governance	0.04	-0.04	-0.09	0.10*	2314
Standardized index of perceived insecurity and drugs	0.06	-0.07	-0.14	0.14	2379
Index of crime	0.09	-0.12	-0.19	0.01**	2379
Other baseline variables					
Index of distance from public goods and services	-0.14	0.14	0.28	0.10	2379
Distance to nearest public transit (meters)	176.55	237.37	60.11	0.17	2379
Distance to nearest cultural center (meters)	92.43	107.20	14.93	0.32	2379
Distance to nearest educational facility (meters)	44.04	77.11	33.24	0.05*	2379
Distance to nearest justice or police center (meters)	556.16	547.61	-5.49	0.46	2379
Distance to nearest religious center (meters)	163.97	168.67	5.21	0.45	2379
Distance to nearest social services (meters)	273.33	328.97	55.26	0.22	2379
Ease to work in sectror for community leaders	1.05	1.30	0.25	0.04**	2379
Area of sector (square meters)	30411.18	29166.02	-1278.39	0.26	2379
Block present in 1970	0.50	0.44	-0.06	0.26	2379
Multidimentional Poverty Index (2018)	14.36	17.30	2.77	0.11	2237
Total population (2018)	2737.45	1583.38	-1171.34	0.25	2379
Percent of women (2018)	52.34	52.31	-0.05	0.47	2379
Percent of population aged 0 to 14 (2018)	18.90	19.47	0.59	0.29	2379
Percent of population aged 15 to 34 (2018)	36.01	37.44	1.46	0.08*	2379
Percent of population who was born on another municipality (2018)	36.24	39.12	2.83	0.07*	2379
Percent of population who recently migrated (2018)	4.24	5.13	0.89	0.06*	2379
Schooling rate (2018)	0.89	0.89	-0.00	0.41	2379
Unemployment rate (2018)	0.11	0.11	0.00	0.32	2379
Median age (2018)	33.77	32.69	-1.11	0.15	2379
Percent of houses with water services (2018)	0.86	0.87	0.01	0.31	2379
Percent of houses with internet services (2018)	0.49	0.48	-0.01	0.36	2379
Percent of houses with electricity (2018)	0.87	0.87	0.00	0.49	2379
Percent of houses with trash collection (2018)	0.87	0.87	-0.00	0.47	2379
Percent of houses with gas services (2018)	0.56	0.59	0.03	0.29	2379
Percent of houses with sewage (2018)	0.86	0.86	-0.00	0.47	2379
Distance to the respective razon headquarters (100 meters)	17.38	19.62	3.50	0.28	1871
Endline survey respondent demographics					
Female	0.66	0.68	0.02	0.18	2379
Respondent age between 18 and 25	0.19	0.19	-0.00	0.50	2379
Respondent age between 26 and 40	0.29	0.31	0.01	0.26	2379
Respondent age between 41 and 64	0.39	0.37	-0.01	0.27	2379
Respondent is business owner	0.20	0.20	0.00	0.47	2379

Notes: This table reports treatment and control group means and a test of balance for the covariates used to match treatment and control sectors (the first five variables) and for some of the covariates selected by the lasso method as prognostic of endline absolute state governance. Regression differences come from an OLS regression of each covariate on an indicator for treatment, calculated at the individual survey level, clustering standard errors at the sector level.

Table D.2: Randomization balance within prespecified subgroups

	High relative state gov.				Low relative state gov.			
	Means		Regression Difference		Means		Regression Difference	
Covariate	Control	Treated	Coeff	RI p-value	Control	Treated	Coeff	RI p-value
Baseline indices used for matching (standardized)								
Standardized index of frequency of combo visibility	0.02	-0.03	0.10	0.35	0.02	-0.03	-0.20	0.11
Standardized values of relative state-combo governance	0.04	-0.04	-0.14	0.08*	0.04	-0.04	-0.05	0.33
Standardized index of perceived insecurity and drugs	0.06	-0.07	-0.11	0.34	0.06	-0.07	-0.16	0.11
Index of crime	0.09	-0.12	-0.21	0.05**	0.09	-0.12	-0.17	0.05*
Other baseline variables								
Index of distance from public goods and services	-0.14	0.14	0.17	0.26	-0.14	0.14	0.39	0.15
Distance to nearest public transit (meters)	176.55	237.37	1.97	0.49	176.55	237.37	116.85	0.12
Distance to nearest cultural center (meters)	92.43	107.20	22.48	0.32	92.43	107.20	7.57	0.44
Distance to nearest educational facility (meters)	44.04	77.11	25.04	0.13	44.04	77.11	41.24	0.12
Distance to nearest justice or police center (meters)	556.16	547.61	-20.16	0.45	556.16	547.61	8.82	0.50
Distance to nearest religious center (meters)	163.97	168.67	-23.45	0.33	163.97	168.67	33.17	0.33
Distance to nearest social services (meters)	273.33	328.97	130.76	0.07*	273.33	328.97	-18.42	0.39
Ease to work in sectror for community leaders	1.05	1.30	0.14	0.21	1.05	1.30	0.34	0.10*
Area of sector (square meters)	30411.18	29166.02	3245.19	0.10	30411.18	29166.02	-5692.25	0.03**
Block present in 1970	0.50	0.44	-0.03	0.42	0.50	0.44	-0.10	0.24
Multidimentional Poverty Index (2018)	14.36	17.30	3.85	0.13	14.36	17.30	1.74	0.28
Total population (2018)	2737.45	1583.38	-2149.97	0.28	2737.45	1583.38	-216.44	0.38
Percent of women (2018)	52.34	52.31	0.75	0.19	52.34	52.31	-0.83	0.15
Percent of population aged 0 to 14 (2018)	18.90	19.47	1.13	0.13	18.90	19.47	0.07	0.48
Percent of population aged 15 to 34 (2018)	36.01	37.44	1.38	0.20	36.01	37.44	1.55	0.12
Percent of population who was born on another municipality (2018)	36.24	39.12	1.11	0.34	36.24	39.12	4.50	0.05**
Percent of population who recently migrated (2018)	4.24	5.13	0.50	0.27	4.24	5.13	1.26	0.07*
Schooling rate (2018)	0.89	0.89	0.03	0.16	0.89	0.89	-0.04	0.10
Unemployment rate (2018)	0.11	0.11	0.00	0.42	0.11	0.11	0.01	0.33
Median age (2018)	33.77	32.69	-1.18	0.15	33.77	32.69	-1.03	0.28
Percent of houses with water services (2018)	0.86	0.87	0.02	0.23	0.86	0.87	0.00	0.46
Percent of houses with internet services (2018)	0.49	0.48	-0.01	0.40	0.49	0.48	-0.02	0.40
Percent of houses with electricity (2018)	0.87	0.87	0.01	0.38	0.87	0.87	-0.01	0.43
Percent of houses with trash collection (2018)	0.87	0.87	0.01	0.33	0.87	0.87	-0.01	0.37
Percent of houses with gas services (2018)	0.56	0.59	0.05	0.27	0.56	0.59	0.01	0.42
Percent of houses with sewage (2018)	0.86	0.86	0.01	0.30	0.86	0.86	-0.01	0.34
Distance to the respective razon headquarters (100 meters)	17.38	19.62	-6.56	0.08*	17.38	19.62	13.12	0.15
Endline survey respondent demographics								
Female	0.66	0.68	0.02	0.27	0.66	0.68	0.03	0.24
Respondent age between 18 and 25	0.19	0.19	0.00	0.41	0.19	0.19	-0.00	0.42
Respondent age between 26 and 40	0.29	0.31	-0.02	0.21	0.29	0.31	0.05	0.08*
Respondent age between 41 and 64	0.39	0.37	0.03	0.19	0.39	0.37	-0.06	0.05*
Respondent is business owner	0.20	0.20	-0.00	0.12	0.20	0.20	0.01	0.10

Notes: This table reports treatment and control group means and a test of balance for all covariates in Table D.1, but does so within the two prespecified subgroups: above and below median baseline relative state governance.

D.3 Accounting for spillovers

We can use our representative city-wide survey to test for spillover effects into non-treated areas, by comparing blocks close to treatment sectors to those close to control sectors, as seen in Appendix Table D.3. We cannot exclude the possibility that there was a minor reduction of service in control sectors, but this should not affect the validity of estimated treatment effects. From an ethical perspective, elected officials undertook the decision to intensify services in the 40 sectors in order to inform future policy.

Note that this approach cannot account for non-Euclidean spillovers. Though we have no reason to believe that the control sectors are more connected in any way to the treatment sectors than any other set of streets in the city, so non-Euclidean spillovers into control sectors seems unlikely.

Table D.3: Estimating treatment spillovers onto blocks within a 250 meter radius

	Treatment Estimate (1)	P-value (2)	0m-250m Spillover Estimate (3)	P-value (4)
Relative State Governance Index	-0.031	0.121	-0.067	0.919
State Governance Index (0-1)	-0.015	0.232	-0.030	0.946
Combo Governance Index (0-1)	0.014	0.378	0.031	0.869
Relative State Legitimacy Index	0.006	0.889	-0.051	0.706
State Legitimacy Index (0-1)	0.011	0.341	-0.011	0.583
Combo Legitimacy Index (0-1)	0.006	0.776	0.036	0.847

Notes: Our sample includes 6977 survey respondents, including 2,379 in the experimental sectors and 4,598 on blocks from the representative city survey. The tale reports treatment estimates along with an indicator for blocks in the experimental sectors and an indicator for blocks within 250 meters of a treated sector. Spillovers in a dense network of blocks can lead to fuzzy clustering, where clusters do not conform to defined areas. Hence we use randomization inference to estimate exact p-values under the sharp null of no treatment effect for any unit, correcting estimates for fuzzy clustering. To address systematic exposure to spillovers due to the geographic distribution, we weight each observation by the inverse probability of each treatment category: treated, <250 meters, and >250 meters.

D.4 Measurement error correlated with treatment

Our governance and legitimacy measures are self-reported survey data, and hence subject to potential response bias. The fact that we see similar results in administrative police crime and call data reduces this concern somewhat. But there are several additional reasons to believe that measurement error is low, or at least not correlated with treatment status.

First, we do not believe that the presence of combos—a familiar and historical part of everyday life in our sectors—significantly distorted responses. We refined survey questions after dozens of qualitative interviews, fine-tuning language, questions, and approach to elicit truthful answers. For data collection, we used an independent survey firm that already conducted annual security surveys to avoid any connection with the intervention, and to minimize experimenter demand effects. They conducted all interviews anonymously and in private, typically indoors. In the context of a secret interview, we believe most respondents answered questions freely and truthfully. Three analyses are consistent with this conclusion.

Second, we can compare our approach against prior efforts. The city has run surveys in the past on "security fees" paid to the combo. City-wide, 19% of our business respondents and 7% of residents report making payments, with negligible non-response. A city survey conducted earlier in the same year reported a 3% payment rate, with 80% non-response. This suggests our approach was actually more successful in eliciting honest responses.

Third, for our results to be spurious would require a very specific pattern of misreporting. Residents would need to systematically under-report state governance or overstate combo governance only in the treatment sectors that had low initial government presence—in essence, the reverse of normal experimenter demand.

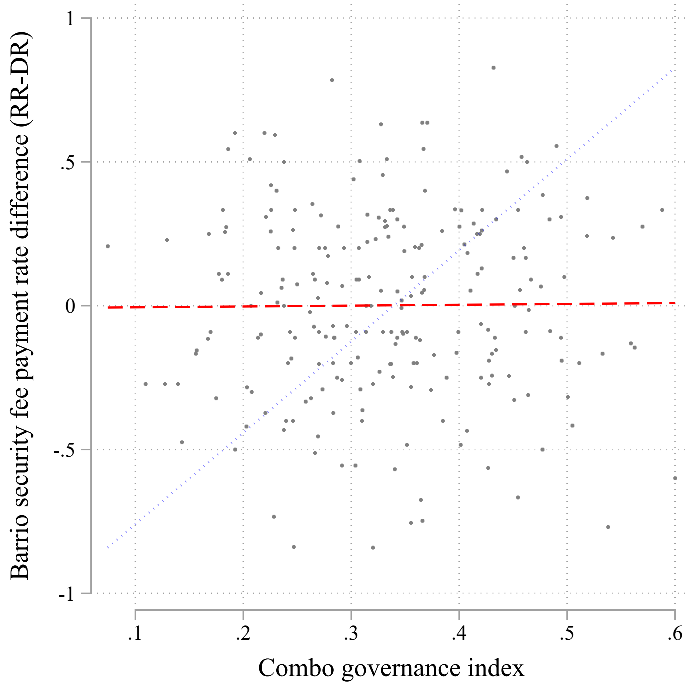
Fourth, we used a survey experiment to assess under-reporting in security fee payment—a measure our qualitative work deemed as one of the most sensitive questions on gang-related activities. We asked some respondents directly whether they paid (Direct response, or DR); others we used a randomized-response (RR) technique, where they privately flipped a coin and responded to the question honestly or not depending on the flip. In other contexts, this method has detected under-reporting of sensitive behaviors.

We see little differences in payment rates between the approaches, suggesting people did not misreport this topic. Randomized response elicited an extortion rate of 22.6% from businesses and 6% from households, compared to 19.4% and 7.8% when directly asked. The differences run in opposite directions and are not statistically significant.

We also see no correlation between assignment to treatment and a RR–DR difference. On average, across all respondents, randomized response results in 4 percentage points higher vacuna payments (not statistically significant). The treatment effect on this RR–DR difference is -0.05, with a standard error of 0.063 (p=0.430).

Figure D.2 also calculates the difference between the RR and DR methods at the barrio level, and plot this difference against combo governance levels. A simple regression line is relatively flat at zero, indicating that misreporting is no more or less common in areas where the combos are more involved in daily life, and hence where legitimacy or fear could potentially have influenced under-reporting.

Figure D.2: Difference between randomized response (RR) and direct response (DR) to survey questions on combo “security fee” payment



Notes: This figure plots the difference between the RR and DR responses to the survey question on extortion against combo governance. Each point represents a barrio average from the 2019 representative city-wide survey. The figure also plots the 45-degree line and a fitted regression line.

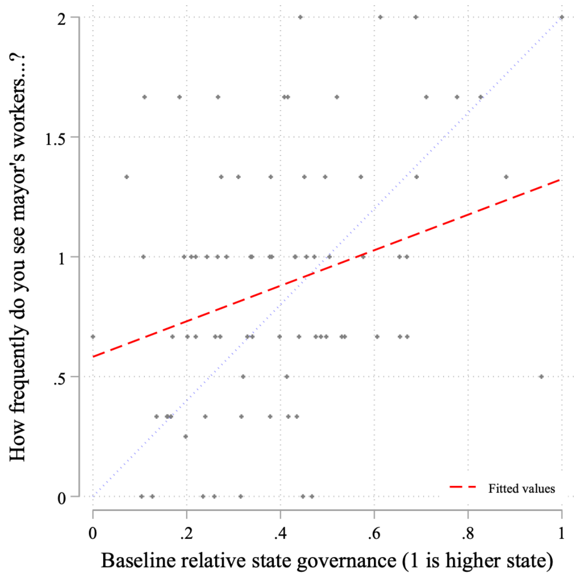
D.5 Identification concerns with heterogeneity analysis

As noted in Section 5.3, a fundamental identification issue with any heterogeneity analysis is that it rests on an assumption of unconfoundedness. Typically, the heterogeneity term is not randomly assigned and may be confounded with other unit characteristics. In our case, we need to ask: does baseline relative state governance drive differing results in the two subgroups, or is it some other sector characteristic that is simply correlated with relative state governance?

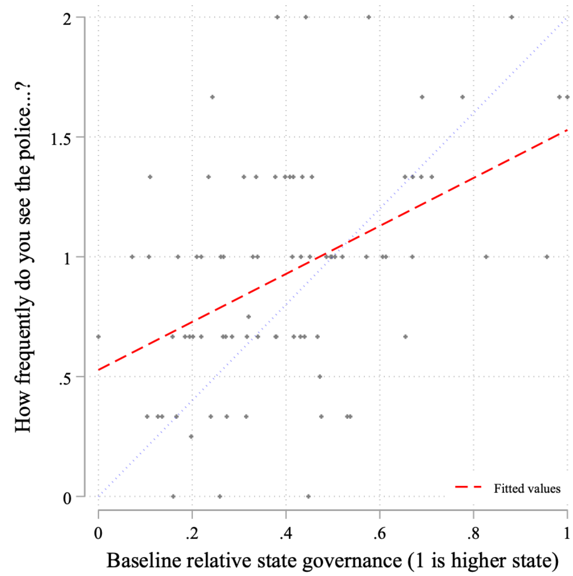
A starting point is to look at observable correlates, and we are fortunate to have a wide range of baseline data on each sector. Table D.4 reports the results of a regression of continuous baseline relative state governance on other baseline covariates at the sector level. The dependent variable is on a [0,1] scale, and is the average of responses across

Figure D.3: How baseline relative state governance correlates with baseline presence of state actors

(a) How frequently mayor's office is observed in community



(b) How frequently police are observed in community



Notes: Both questions.

leaders/officials for that sector. A score closer to 0 implies that the combo mainly responds to governance issues, a 1 implies it is the state, and 0.5 corresponds to both equally.

Two of the largest and most statistically significant correlation are leader survey measures of the frequency with which police and Alcaldía workers are seen in the neighborhood (especially police). We plot these bivariate relationships in Figure D.3, and see that the relationship holds for Mayor's visibility when police are not included (suggesting the two are highly correlated).

Interestingly, after controlling for these measures of observed presence, an index of distance from various state headquarters and infrastructure (such as schools) and the percent of houses with utility connections are not correlated with baseline relative state governance in Table D.4.

Another way to see if the baseline relative state governance measure is associated with absolute state governance is to look at the 40 control sectors only and compare baseline rela-

Table D.4: Pretreatment correlates of baseline relative state governance

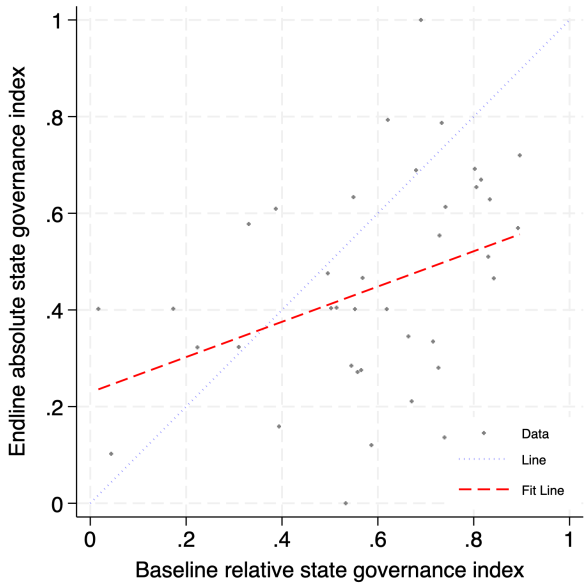
	(1) No FE
Frequency see mayor's workers on street (0-1)	0.256
	(0.205)
Frequency see police on street (0-1)	0.285**
	(0.111)
Frequency see combo on street (0-1)	0.188***
	(0.0678)
Standardized index of perceived insecurity and drugs	0.00630
	(0.0257)
Index of crime (standardized)	0.0108
	(0.0176)
Index of distance from public goods and services (standardized)	0.0126
	(0.0221)
Proportion of houses with utility connections (standardized)	0.0145
	(0.0291)
Ease to work in sector for community leaders (0-3)	-0.0303
	(0.0333)
Area of sector in square meters (standardized)	-0.0459**
	(0.0229)
Block present in 1970 (standardized)	-0.0449*
	(0.0227)
Total population (standardized)	0.675
	(0.915)
Total women (standardized)	-0.462
	(0.487)
Total population aged 0 to 14 (standardized)	0.202
	(0.315)
Total population aged 15 to 34 (standardized)	-0.412
	(0.599)
Percent of population who was born on another municipality (2018) (standardized)	-0.0327
	(0.0223)
Percent of population who recently migrated (2018) (standardized)	-0.0353
	(0.0243)
Schooling rate (2018) (standardized)	0.0204
	(0.0159)
Unemployment rate (2018) (standardized)	-0.00945
	(0.0261)
Median age (2018) (standardized)	-0.0673
	(0.0757)
Observations	80

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: We regress our continuous heterogeneity measure (baseline relative state governance) on other baseline covariates at the sector level, with and without block-pair fixed effects.

Figure D.4: Relative vs absolute state governance, 40 control sectors



tive state governance (according to leaders) to endline absolute state governance (according to citizens), as in Figure D.4. The two measures are highly correlated.

Together these analyses suggest that our baseline relative heterogeneity measure may be a reasonable proxy for state penetration—both of police and the Alcaldía.

Interestingly, one of the other large and significant correlations is with the frequency combo members are seen on the street. It too is *positively* correlated with low baseline relative state governance. This is consistent with the fact that we do not see a negative correlation between state and combo governance overall (Section 2). As discussed in the paper, other evidence suggests that combos may govern more in response to high police presence, to guard drug revenues.

As a result, we think our baseline relative state governance measure may be reasonably interpreted as a proxy for the quality of community governance generally. Of course, there are many other potential unobservables and so we must take this interpretation with caution.

Few other variables are statistically significant. That said, larger sectors and older sectors have somewhat lower baseline relative state governance. More populous and more youthful sectors, and those with fewer migrants have higher baseline relative state governance (not statistically significant, though the coefficients are large). Speculatively, this would be consistent with a correlation in social cohesion/trust and baseline relative state governance. This is in turn consistent with one of our hypotheses that low governance is endogenous to social capital.

D.6 Additional intervention details

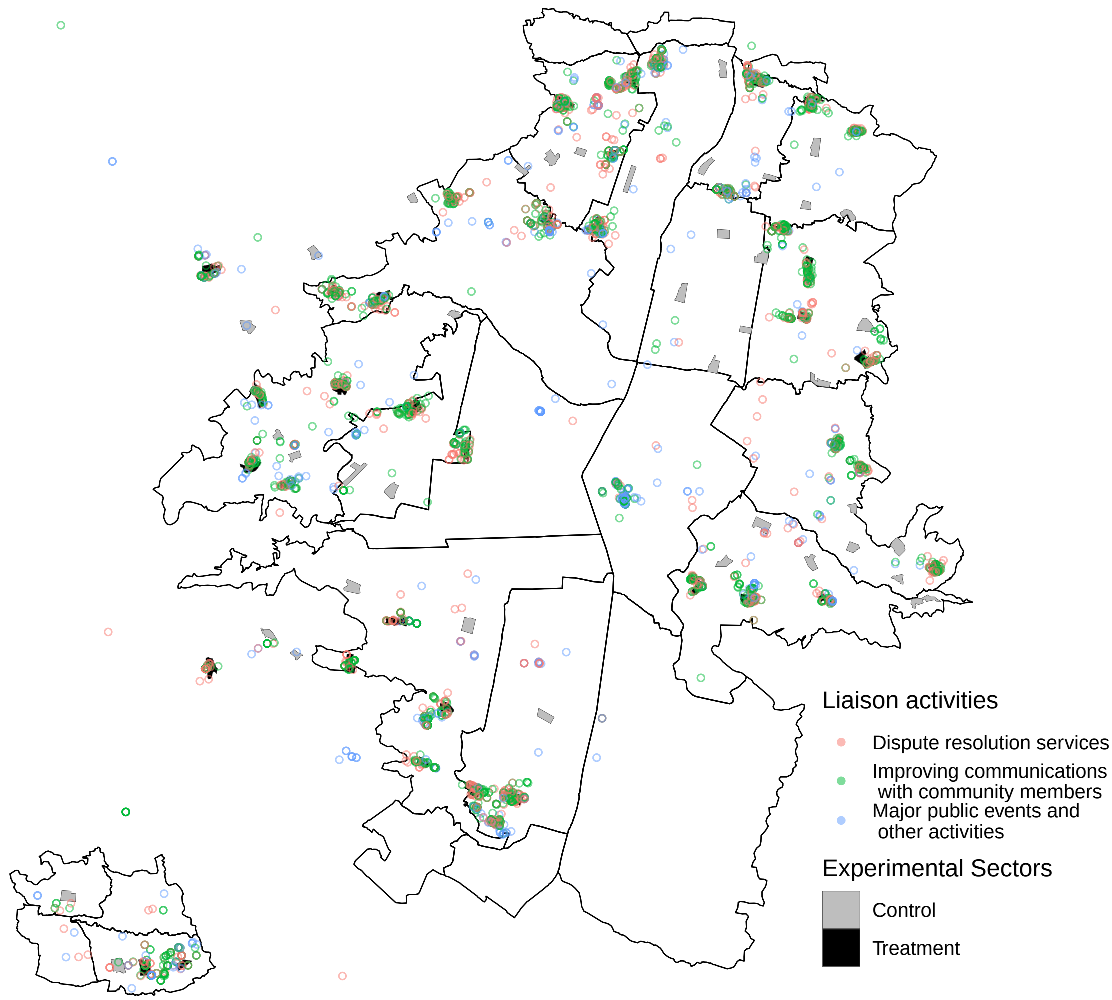
Liaisons were instructed by the Secretariat of Security to log the activities they undertook in their sectors. Most of these activities happened within the sector, as expected given the nature of the intervention. In some cases, however, liaisons implemented and logged activities outside the perimeter of their sector. This happened because of logistical requirements: activities that took place in a nearby soccer field, at dispute resolution offices, or at the city administration headquarters, among others. Figure D.5 depicts the logged activities by experimental sector, geocoded. Table D.5 presents the treatment effect of logged activities in treatment sectors.

Table D.5: Count of officially-logged liaison activities per sector

			Het. by baseline rel. gov.			
		ATE	Above median	Below median	Diff.	
	Control Mean (SD)	Estimate [p-value]	Estimate [p-value]	Estimate [p-value]	Estimate [p-value]	N
	(1)	(2)	(3)	(4)	(5)	(6)
N. activities in 125m sector buffer area per month	0.188	6.927***	7.091***	6.780***	0.311	80
	(0.475)	[0.000]	[0.000]	[0.000]	[0.901]	80
N. activities per month	0.000	9.853***	9.896***	9.814***	0.311	80
	(0.000)	[0.000]	[0.000]	[0.000]	[0.974]	80

Notes: To illustrate treatment compliance, and spillovers of liaison activities into control sectors, this table reports summary statistics and treatment effects for all activities officially logged by liaisons. We examine the count of all activities within the experimental sector itself, as well as within a 125 meter buffer. Note that the activities of other municipal employees are not logged, and so this is an incomplete measure of city staff activities. Average treatment effects and treatment heterogeneity are calculated using the same approach as in Table 3. Note that only households and not businesses were surveyed on legitimacy.

Figure D.5: Alcaldía and liaison activities by experimental sector



Notes: Liaisons were instructed to report their major events and referrals, including the location. This figure depicts all major liaison activities conducted, overlaid atop a map of treatment and control sectors. We group the interventions into three subgroups based on their goals. Treatment sectors are mostly obscured by the density of activities, while control communities receive relatively few.

E Supplemental tables

Table E.1: Program impacts on legitimacy components

Dependent variable	Control Mean (SD) (1)	ATE: Estimate (CRSE p) [RI p] (2)	Het. by baseline rel. gov.: Above median: Estimate (CRSE p) [RI p] (3)	Het. by baseline rel. gov.: Below median: Estimate (CRSE p) [RI p] (4)	Het. by baseline rel. gov.: Diff.: Estimate (CRSE p) [RI p] (5)	N
Police legitimacy index	0.57	0.010	0.038**	-0.022*	0.059***	1,906
	(0.23)	(0.275)	(0.002)	(0.029)	(0.000)	
		[0.220]	[0.025]	[0.079]	[0.003]	
How much do you trust the police	0.56	0.006	0.039**	-0.032*	0.071***	1,900
	(0.34)	(0.630)	(0.008)	(0.039)	(0.001)	
		[0.391]	[0.045]	[0.063]	[0.008]	
How fair is the police	0.57	-0.002	0.013	-0.019	0.031	1,838
	(0.30)	(0.824)	(0.346)	(0.217)	(0.108)	
		[0.434]	[0.276]	[0.196]	[0.146]	
How do you rate the police	0.59	0.011	0.037**	-0.018	0.056**	1,871
	(0.24)	(0.273)	(0.004)	(0.110)	(0.001)	
		[0.213]	[0.033]	[0.140]	[0.013]	
How would your neighbors rate the police	0.59	0.019*	0.041**	-0.005	0.047**	1,771
	(0.26)	(0.043)	(0.001)	(0.647)	(0.006)	
		[0.072]	[0.022]	[0.375]	[0.039]	
How much do your neighbors trust the police	0.57	0.016	0.060**	-0.034*	0.094***	1,780
	(0.32)	(0.191)	(0.000)	(0.010)	(0.000)	
		[0.186]	[0.019]	[0.053]	[0.000]	
Mayor legitimacy index	0.57	0.014	0.028*	-0.002	0.030	1,906
	(0.23)	(0.109)	(0.021)	(0.847)	(0.087)	
		[0.188]	[0.080]	[0.428]	[0.130]	
How much do you trust the mayoral staff	0.57	0.007	0.023	-0.010	0.033	1,881
	(0.33)	(0.533)	(0.156)	(0.533)	(0.166)	
		[0.383]	[0.191]	[0.341]	[0.186]	
How fair is the mayoral staff	0.53	0.008	0.024	-0.010	0.034	1,776
	(0.31)	(0.486)	(0.131)	(0.549)	(0.146)	
		[0.314]	[0.172]	[0.326]	[0.160]	
How do you rate the mayoral staff	0.61	0.005	0.020	-0.012	0.032	1,857
	(0.25)	(0.632)	(0.132)	(0.400)	(0.092)	
		[0.390]	[0.170]	[0.285]	[0.121]	
How would your neighbors rate the mayoral staff	0.59	0.022	0.024	0.019	0.006	1,708
	(0.27)	(0.037)	(0.101)	(0.197)	(0.776)	
		[0.105]	[0.152]	[0.249]	[0.453]	
How much do your neighbors trust the mayoral staff	0.55	0.033**	0.047*	0.018	0.029	1,761
	(0.32)	(0.003)	(0.014)	(0.138)	(0.220)	
		[0.040]	[0.060]	[0.225]	[0.221]	
Combo legitimacy index	0.44	-0.002	-0.015	0.012	-0.027	1,845
	(0.28)	(0.877)	(0.394)	(0.613)	(0.357)	
		[0.461]	[0.278]	[0.363]	[0.259]	
How much do you trust the combo	0.36	0.006	-0.006	0.019	-0.025	1,822
	(0.36)	(0.742)	(0.720)	(0.516)	(0.447)	
		[0.408]	[0.392]	[0.316]	[0.284]	
How fair is the combo	0.41	0.005	-0.022	0.035	-0.057	1,689
	(0.34)	(0.760)	(0.352)	(0.150)	(0.103)	
		[0.443]	[0.286]	[0.166]	[0.124]	
How do you rate the combo	0.50	0.003	-0.010	0.018	-0.027	1,642
	(0.27)	(0.794)	(0.537)	(0.420)	(0.315)	
		[0.419]	[0.356]	[0.298]	[0.243]	
How much do your neighbors trust the combo	0.51	-0.009	-0.027	0.011	-0.038	1,618
	(0.30)	(0.540)	(0.130)	(0.608)	(0.175)	
		[0.330]	[0.175]	[0.372]	[0.196]	
How would your neighbors rate the combo	0.48	-0.008	-0.021	0.007	-0.027	1,671
	(0.36)	(0.677)	(0.425)	(0.788)	(0.447)	
		[0.381]	[0.277]	[0.410]	[0.291]	

Notes: The table reports summary statistics and treatment effects for 5 survey-based measures of legitimacy per actor, plus a summary index for the 5 questions. Each row is a different dependent variable. Each row is a different dependent variable. Average treatment effects and treatment heterogeneity are calculated using the same approach as in Table 3. Note that only households and not businesses were surveyed on legitimacy.

Table E.2: Program impacts on state governance components

			Het. by baseline rel. gov.			
		ATE	Above median	Below median	Diff.	
Dependent variable	Control Mean (SD)	Estimate (CRSE p) [RI p]	Estimate (CRSE p) [RI p]	Estimate (CRSE p) [RI p]	Estimate (CRSE p) [RI p]	N
	(1)	(2)	(3)	(4)	(5)	
Relative state governance index (less police related)	0.09	-0.021	-0.015	-0.026	0.011	2,279
	(0.31)	(0.074)	(0.343)	(0.086)	(0.611)	
		[0.118]	[0.279]	[0.147]	[0.368]	
HH: Someone is making noise	0.26	-0.019	-0.021	-0.016	-0.005	1,747
	(0.42)	(0.234)	(0.224)	(0.545)	(0.866)	
		[0.238]	[0.208]	[0.370]	[0.458]	
HH: Home improvements affect neighbors	0.14	0.004	0.011	-0.004	0.016	1,567
	(0.44)	(0.847)	(0.630)	(0.880)	(0.663)	
		[0.467]	[0.393]	[0.459]	[0.386]	
HH: There is domestic violence	0.15	-0.004	0.028	-0.037	0.065	1,559
	(0.45)	(0.810)	(0.256)	(0.147)	(0.067)	
		[0.431]	[0.226]	[0.204]	[0.124]	
HH: Two drunks fight on the street	0.13	-0.004	0.004	-0.014	0.018	1,645
	(0.45)	(0.826)	(0.885)	(0.608)	(0.659)	
		[0.450]	[0.458]	[0.388]	[0.379]	
Biz: Someone disturbs a business	0.16	-0.087*	-0.050	-0.128**	0.078	382
	(0.50)	(0.069)	(0.515)	(0.015)	(0.397)	
		[0.091]	[0.330]	[0.031]	[0.270]	
HH: People smoking marijuana near children	0.03	0.003	0.022	-0.017	0.039	1,682
	(0.40)	(0.868)	(0.381)	(0.537)	(0.303)	
		[0.473]	[0.280]	[0.352]	[0.248]	
HH: Kids fight on the street	-0.03	-0.017	-0.005	-0.028	0.023	1,552
	(0.41)	(0.339)	(0.841)	(0.263)	(0.498)	
		[0.254]	[0.447]	[0.206]	[0.312]	
Biz: Someone does not want to pay a debt	-0.05	-0.006	0.011	-0.024	0.036	370
	(0.33)	(0.817)	(0.728)	(0.558)	(0.475)	
		[0.402]	[0.409]	[0.309]	[0.287]	
HH: Someone refuses to pay a big debt	-0.20	-0.032*	-0.005	-0.058**	0.053	1,434
	(0.45)	(0.092)	(0.831)	(0.008)	(0.063)	
		[0.094]	[0.438]	[0.023]	[0.141]	
Relative state governance index (more police related)	0.02	-0.032	-0.016	-0.049	0.033	2,239
	(0.40)	(0.076)	(0.541)	(0.039)	(0.342)	
		[0.130]	[0.360]	[0.103]	[0.265]	
Biz: You have to react to a robbery	0.12	-0.097**	-0.133**	-0.057	-0.075	372
	(0.48)	(0.026)	(0.014)	(0.382)	(0.355)	
		[0.044]	[0.050]	[0.248]	[0.263]	
Biz: It is necessary to prevent a theft	0.08	-0.078*	-0.069	-0.088	0.019	396
	(0.52)	(0.075)	(0.302)	(0.076)	(0.813)	
		[0.093]	[0.232]	[0.106]	[0.432]	
Biz: Businesses in this sector are robbed	0.07	-0.072*	-0.075	-0.069	-0.006	362
	(0.50)	(0.064)	(0.154)	(0.227)	(0.943)	
		[0.099]	[0.170]	[0.206]	[0.487]	
HH: A car or motorbike is stolen	-0.01	0.020	0.053	-0.017	0.070	1,557
	(0.45)	(0.346)	(0.074)	(0.578)	(0.097)	
		[0.284]	[0.133]	[0.358]	[0.149]	
HH: Someone is threatening someone else	-0.01	-0.026	0.013	-0.065	0.078	1,589
	(0.45)	(0.244)	(0.600)	(0.062)	(0.063)	
		[0.209]	[0.365]	[0.105]	[0.104]	
HH: You have to react to a robbery	-0.02	-0.017	0.012	-0.048	0.060	1,635
	(0.47)	(0.445)	(0.692)	(0.116)	(0.153)	
		[0.296]	[0.399]	[0.155]	[0.184]	
HH: Someone is mugged on the street	-0.05	0.020	0.044	-0.005	0.049	1,569
	(0.42)	(0.354)	(0.191)	(0.856)	(0.260)	
		[0.281]	[0.204]	[0.443]	[0.231]	
HH: It is necessary to prevent a theft	-0.04	-0.007	0.025	-0.042	0.067	1,692
	(0.48)	(0.759)	(0.415)	(0.183)	(0.116)	
		[0.404]	[0.314]	[0.197]	[0.153]	

*Notes:*The table reports summary statistics and treatment effects for the 17 components of the governance index in Table 3. We create sub-indexes for what our qualitative work suggests are more and less police-related forms of governance. Each row is a different dependent variable. Average treatment effects and treatment heterogeneity are calculated using the same approach as in Table 3. Note that both households and businesses were surveyed on governance.

Table E.3: Impacts of treatment on survey measures of police, mayoral, and combo efficacy

Dependent variable	Control Mean (SD) (1)	ATE Estimate (CRSE p) [RI p] (2)	Het. by baseline rel. gov. Above median Estimate (CRSE p) [RI p] (3)	Het. by baseline rel. gov. Below median Estimate (CRSE p) [RI p] (4)	Het. by baseline rel. gov. Diff. Estimate (CRSE p) [RI p] (5)	N
Police efficacy index	0.55	-0.006	-0.001	-0.011	0.009	1,906
	(0.21)	(0.533)	(0.917)	(0.279)	(0.578)	
		[0.333]	[0.478]	[0.225]	[0.359]	
How easy is it to contact the police	0.54	-0.008	-0.002	-0.014	0.012	1,869
	(0.29)	(0.429)	(0.897)	(0.116)	(0.504)	
		[0.305]	[0.490]	[0.146]	[0.299]	
Perceived value of the police	0.71	0.003	-0.003	0.011	-0.014	1,864
	(0.25)	(0.734)	(0.846)	(0.248)	(0.467)	
		[0.394]	[0.433]	[0.237]	[0.321]	
How fast is the police	0.42	-0.010	0.006	-0.029	0.035	1,868
	(0.34)	(0.488)	(0.717)	(0.198)	(0.180)	
		[0.308]	[0.423]	[0.168]	[0.201]	
Mayoral staff efficacy index	0.45	-0.008	-0.001	-0.016	0.015	1,893
	(0.20)	(0.359)	(0.917)	(0.108)	(0.369)	
		[0.263]	[0.462]	[0.144]	[0.283]	
How easy is it to contact mayoral staff	0.35	-0.000	0.002	-0.002	0.004	1,745
	(0.30)	(0.985)	(0.928)	(0.871)	(0.860)	
		[0.478]	[0.466]	[0.488]	[0.460]	
Perceived value of the mayoral staff	0.66	-0.010	0.005	-0.028**	0.033	1,823
	(0.26)	(0.284)	(0.753)	(0.001)	(0.057)	
		[0.217]	[0.427]	[0.012]	[0.104]	
How fast is the mayoral staff	0.34	-0.006	0.008	-0.021	0.029	1,792
	(0.31)	(0.634)	(0.633)	(0.172)	(0.189)	
		[0.378]	[0.381]	[0.190]	[0.202]	
Combo efficacy index	0.55	-0.001	-0.002	0.001	-0.003	1,790
	(0.24)	(0.934)	(0.895)	(0.975)	(0.906)	
		[0.469]	[0.478]	[0.496]	[0.486]	
How easy is it to contact the combo	0.59	0.014	0.024	0.002	0.022	1,649
	(0.31)	(0.386)	(0.169)	(0.923)	(0.469)	
		[0.289]	[0.213]	[0.478]	[0.307]	
Perceived value of the combo	0.52	-0.012	-0.032	0.008	-0.040	1,706
	(0.32)	(0.390)	(0.064)	(0.720)	(0.157)	
		[0.294]	[0.117]	[0.406]	[0.182]	
How fast is the combo	0.56	0.007	0.005	0.009	-0.004	1,589
	(0.36)	(0.671)	(0.848)	(0.627)	(0.908)	
		[0.383]	[0.438]	[0.360]	[0.469]	

Notes: This table reports summary statistics and treatment effects on 3 survey-based measures of efficacy per actor, plus a summary index for the three questions. Each row is a different dependent variable. Average treatment effects and treatment heterogeneity are calculated using the same approach as in Table 3. Note that only households and not businesses were surveyed on efficacy.

Table E.4: Robustness table

	Control Mean	N	Main spec.	No controls	With respondent demog.	Lasso controls
Average treatment effect						
Family index of all indices (z-score)	-0.11	80	0.246	0.222	0.317*	0.135
			(0.133)	(0.134)	(0.092)	(0.253)
Relative state legitimacy index	0.13	1845	0.016	0.011	0.021	0.041*
			(0.278)	(0.332)	(0.219)	(0.059)
Relative state governance index	0.07	2314	-0.025	-0.028	-0.022	-0.017
			(0.124)	(0.109)	(0.143)	(0.214)
Sentence-weighted crime index	0.35	80	-0.061*	-0.063*	-0.034	-0.032
			(0.066)	(0.072)	(0.220)	(0.174)
Security-related emergency calls	135.75	80	-34.969**	-37.375**	-34.274**	-28.387**
			(0.028)	(0.010)	(0.040)	(0.038)
ATE in above median baseline governance						
Family index of all indices (z-score)	-0.11	80	0.665**	0.672***	0.832**	0.520**
			(0.016)	(0.007)	(0.010)	(0.033)
Relative state legitimacy index	0.13	1845	0.050*	0.072	0.057*	0.051*
			(0.084)	(0.137)	(0.058)	(0.068)
Relative state governance index	0.07	2314	-0.018	0.037	-0.016	0.009
			(0.294)	(0.213)	(0.311)	(0.408)
Sentence-weighted crime index	0.35	80	-0.137**	-0.135**	-0.112*	-0.058
			(0.024)	(0.040)	(0.070)	(0.151)
Security-related emergency calls	135.75	80	-63.250**	-64.400***	-66.878***	-41.691*
			(0.011)	(0.006)	(0.009)	(0.056)

Notes: This tables replicates our main results (Column 1) and re-estimates average treatment effects with three alternative control vectors: no added covariates (Column 2); standard covariates plus survey respondent demographics (Column 2), including gender, age, and a resident versus business indicator; and a control vector determined via a lasso regression of the dependent variables on all available pre-treatment covariates (Column 3). We also construct a standardized family index of all four outcomes as a simple unweighted average of the four indexes, themselves standardized before averaging. We calculate p-values through randomization inference.

Table E.5: Program impacts on legitimacy and governance: Average treatment effects and heterogeneity by predicted baseline governance quality

Dependent variable	Control Mean (SD) (1)	ATE Estimate (CRSE p) [RI p] (2)	Het. by baseline pred. gov.: Above median Estimate (CRSE p) [RI p] (3)	Het. by baseline pred. gov.: Below median Estimate (CRSE p) [RI p] (4)	Het. by baseline pred. gov.: Diff. Estimate (CRSE p) [RI p] (5)	N
Relative state legitimacy index	0.13	0.016	0.035	-0.006	0.041	1,845
	(0.32)	(0.392)	(0.195)	(0.807)	(0.268)	
		[0.291]	[0.213]	[0.440]	[0.242]	
State legitimacy index	0.57	0.013	0.020	0.005	0.015	1,906
	(0.21)	(0.090)	(0.049)	(0.659)	(0.335)	
		[0.145]	[0.106]	[0.397]	[0.269]	
Combo legitimacy index	0.44	-0.002	-0.013	0.012	-0.025	1,845
	(0.28)	(0.877)	(0.556)	(0.537)	(0.393)	
		[0.461]	[0.355]	[0.360]	[0.295]	
Relative state governance index	0.07	-0.025	-0.011	-0.035	0.024	2,314
	(0.31)	(0.068)	(0.563)	(0.059)	(0.366)	
		[0.125]	[0.367]	[0.109]	[0.283]	
State governance index	0.41	-0.012	0.021**	-0.026**	0.047***	2,362
	(0.26)	(0.162)	(0.006)	(0.000)	(0.000)	
		[0.186]	[0.044]	[0.015]	[0.003]	
Combo governance index	0.35	0.011	0.028	0.007	0.022	2,316
	(0.28)	(0.389)	(0.068)	(0.732)	(0.378)	
		[0.280]	[0.141]	[0.399]	[0.291]	

Notes: The table reports ITT estimates of program impacts and treatment heterogeneity. Each row is a different dependent variable. We report p-values from cluster robust standard error estimation (CRSE) in parentheses and from randomization inference (RI) in brackets. Both households and businesses were surveyed on governance levels (N=2,379), but only households were surveyed on legitimacy and hence there are fewer observations (N=1,910).

Table E.6: Heterogeneity analysis by quartiles of relative baseline state governance

	Dependent Variable: Relative governance	Dependent Variable: Relative legitimacy	Dependent Variable: Sentence-weighted crime index	Dependent Variable: Security-related emergency calls
	(1)	(2)	(3)	(4)
Program impacts:				
Q1 (0^{th} - 25^{th} quartile baseline rel. gov)	-0.025	-0.071	-0.019	-14.369
	(0.332)	(0.146)	(0.386)	(0.352)
Q2 (25^{th} - 50^{th} quartile baseline rel. gov)	-0.038	0.018	0.058	5.289
	(0.113)	(0.411)	(0.155)	(0.435)
Q3 (50^{th} - 75^{th} quartile baseline rel. gov)	0.022	0.056	-0.086	-57.214*
	(0.291)	(0.113)	(0.182)	(0.071)
Q4 (75^{th} - 100^{th} quartile baseline rel. gov)	-0.061	0.038	-0.194**	-70.493*
	(0.123)	(0.281)	(0.050)	(0.062)
Differences relative to Q1:				
Q2	-0.013	0.089	0.076	19.658
	(0.446)	(0.172)	(0.196)	(0.320)
Q3	0.048	0.127*	-0.067	-42.845
	(0.227)	(0.055)	(0.250)	(0.210)
Q4	-0.035	0.109	-0.175*	-56.124
	(0.331)	(0.110)	(0.092)	(0.182)

Notes: This table replicates the results of Table 3 but partitioning the sample in 4 subgroups (quartiles) as opposed to 2. Here we report program effects on each each subgroup in the first 4 rows, while the last 3 report differences with respect to the lowest governance group. Unfortunately this leaves just 20 sectors per quartile subgroup, making this analysis somewhat underpowered. Both households and businesses were surveyed on governance levels, but only households were surveyed on legitimacy (and hence there are fewer observations).