

WORKING PAPER · NO. 2022-20

State-building on the Margin: An Urban Experiment in Medellín

Christopher Blattman, Gustavo Duncan, Benjamin Lessing, and Santiago Tobon

JUNE 2022

State-building on the Margin: Lessons from an Urban Experiment in Medellín*

Christopher Blattman Gustavo Duncan Benjamin Lessing
Santiago Tobón[†]

June 14, 2022

Abstract

Medellín’s city government wanted to raise its efficacy and legitimacy, especially in neighborhoods with weak state presence and competing armed actors. The city identified 80 such neighborhoods. In half, they intensified non-police street presence tenfold for two years, attempting to improve social services and dispute resolution. Unexpectedly, despite increased attention from street staff, residents lowered their opinion of the state on average. We trace these adverse effects to communities where state presence was initially weakest. Staff in these neighborhoods worked to improve services, but the central administration struggled to deliver on these promises. Where state presence was already established, however, the intervention raised opinions of the government as expected. We hypothesize that it is costlier for states to improve services where it is weak—an incentive for bureaucrats and elected leaders to concentrate state-building efforts in established areas, widening inequality in public service access and local variance in state legitimacy.

*For comments and feedback we thank Oriana Bandiera, Eli Berman, Macartan Humphreys, Raul Sánchez de la Sierra, Jacob Shapiro, Carlos Schmidt-Padilla, Paolo Pinotti, Maria Micaela Sviatschi, Juan F. Vargas, and participants at several seminars and conferences. Innovations for Poverty Action coordinated all research activities. For research assistance we thank Bruno Aravena, David Cerero, Peter Deffebach, Felipe Fajardo, Sebastián Hernández, Sofía Jaramillo, Juan F. Martínez, Juan Pablo Mesa-Mejía, Angie Mondragón, Helena Montoya, José Miguel Pascual, Andrés Preciado, M. Aránzazu Rodríguez-Uribe, Zachary Tausanovitch, Nelson Matta-Colorado and Martín Vanegas. We thank the Secretariat of Security of Medellín for their cooperation, especially the former Secretary of Security Andrés Tobón, as well as Lina Calle and Ana María Corpas. For financial support, we thank the Centro de Estudios sobre Seguridad y Drogas (CESED) of Universidad de los Andes; the Peace and Recovery Program (P&R) at Innovations for Poverty Action (IPA); the PROANTIOQUIA foundation; The National Science Foundation (NSF); the UK Foreign, Commonwealth & Development Office through the Crime and Violence Initiative at J-PAL; and the Economic Development and Institutions Programme (EDI) funded with UK aid from the UK Government, working in partnership with Oxford Policy Management Limited, University of Namur, Paris School of Economics and Aide à la Décision Économique.

[†]Blattman (corresponding author): University of Chicago, Harris School of Public Policy & Department of Political Science blattman@uchicago.edu; Duncan: Universidad EAFIT, Department of Political Science gduncan@eafit.edu.co; Lessing: University of Chicago, Department of Political Science blissing@uchicago.edu; Tobon: Universidad EAFIT, Department of Economics stobonz@eafit.edu.co

1 Introduction

Historically, state building has been a slow process of developing organizational capacity to collect taxes, deliver services, and order society. States commonly share governance functions with traditional leaders and community organizations (Cammett and MacLean, 2014; Van der Windt et al., 2019; Blattman et al., 2014; Henn, 2021). This is unproblematic in many realms, but for things that require coercive capacity—basic order, public security, and justice—non-state actors may undermine state legitimacy and service use (e.g., Berman and Laitin, 2008; Acemoglu et al., 2020; Cammett and MacLean, 2014). This is surely one reason why modern states aim to monopolize the legitimate use of force (Weber, 1946; Tilly, 1990).

This Weberian ideal was rarely achieved in practice throughout much of history, and it remains elusive in the developing world today. In many places, basic security and dispute resolution services are provided by armed groups and criminal gangs. Such “criminal governance” exists in hundreds of cities worldwide (Arias, 2006). Moreover, this criminal rule often overlaps with state rule, or is surrounded by areas of strong government presence (Lessing, 2020). This is puzzling to scholars and governments alike.

The conventional wisdom is that both armed and unarmed actors increase their governance in response to a state that is unable or unwilling to project power. Scholars trace the origins of the Sicilian Mafia, Brazilian and California prison gangs, Congolese warlords, and other groups to the state’s inability to protect production or regulate illegal transactions (Gambetta, 1996; Skaperdas, 2001; Gray, 2003; Arias, 2006; Skarbek, 2011; Acemoglu et al., 2020; Sánchez De La Sierra, 2020). Insurgent groups engage in similar governance efforts to foster civilian support in the territories they control (Berman et al., 2011; Arjona, 2016; Kasfir, 2015). More broadly, a literature on fixing failed states also emphasizes that weak states must fill sovereignty gaps and empower communities to move away from warlord rule (Ghani and Lockhart, 2009; Del Castillo, 2008; Karim, 2020).

What can states do in this situation? Their presence and legitimacy is uneven, even

within their principal cities. Wealthy zones with high levels of state capacity about more marginal zones characterized by weaker state presence and varying degrees of criminal governance. Many states want to shore up their presence and legitimacy in these marginal zones. How to do so?

This paper uses an unusual experiment in Medellín to explore whether expanding state presence can earn the trust and engagement of citizens, and improve the state’s position as the preeminent provider of security. The short-term returns to such investments in state penetration are poorly understood. The costs of hiring bureaucrats, extending public services, and expanding fiscal capacity are often immediate and clear. But what should a mayor expect to receive in terms of efficacy, legitimacy, and public support in the near term?

This is a question governments ask in every society, but it takes on special importance in cities where illegal groups compete for citizen loyalty. In Medellín, for example, most low- and middle-income neighborhoods have a neighborhood gang called a *combo*. In addition to selling drugs, many combos police commercial streets and settle disputes between neighbors for a fee, undermining the state’s monopoly on coercion.

In situations like this, a key question is how returns to investments in state presence depend on initial levels of state presence and legitimacy. On the one hand, in areas with little history of state services, we might expect the first investments to have out-sized impacts. This was our initial hypothesis in Medellín, where residents of the least-served areas initially expressed relief at finally seeing municipal bureaucrats in their neighborhoods.¹ On the other hand, establishing thoroughgoing state governance and legitimacy might require large and sustained investments, especially from a low starting point. For example, investments that prove “too small” could have little or even negative impact on residents’ perceptions of the state, particularly if actions fail to live up to promises.

In 2018, we worked with Medellín’s municipal government, the *Alcaldía*, to identify 80

¹This hypothesis finds support in recent literature suggesting that the returns to government investments in fostering political participation might be highest in places where the state is weakest, following tax collection efforts (Weigel, 2020).

areas in need of more state governance. We show they are fairly representative of poor- and middle-income neighborhoods in the city. All 80 had a significant combo presence, though the degree of criminal governance varied widely across the sample. We randomly selected 40 of these neighborhoods to receive intensive city services for two years. To our knowledge, it is the first randomized trial of a gang-level intervention of any kind in the world.²

The intervention had two main aspects. First, the Alcaldía created a special task force to ensure that any needs identified in these communities would get priority attention in the city’s many service agencies. Second, in each neighborhood they hired a full-time “liaison”—a street-level bureaucrat whose job was to: rejuvenate community government organizations; advertise and link people to government agencies; identify public service needs (such as garbage pickup or poor playgrounds) for the task force to fix; and, in the case of serious disputes, connect disputants to professional mediators from the city. There was no change in state criminal justice attention.

The premise of the intervention was simple: by improving public-service delivery, providing non-criminal alternatives for dispute resolution, and strengthening the ability of local groups to identify problems and solutions to everyday community problems, the Alcaldía could increase its relative legitimacy and citizen use of its services.

This was a highly intensive increase in state presence. These were small neighborhoods, with about 200–600 households each, and we estimate they received a tenfold increase in street-level attention to problems for about 20 months. We monitored liaisons closely, and confirmed a high level of compliance.

We prespecified two main outcomes: relative state-versus-combo service access (“relative state governance”) and relative state-versus-combo public trust and legitimacy (“relative state legitimacy”). Both are indices capturing people’s opinions on a set of different situations

²A recent review of interventions outside the OECD found just four small case studies (Higginson et al., 2015). Even within the US, most research has focused mainly on the determinants of gang affiliation and risk factors (e.g., Craig et al., 2002; Cureton, 1999; Curry et al., 2002; Decker and Curry, 2000). There is, however, some empirical evidence on specific programs although none includes large experimental trials (e.g., Esbensen et al., 2001; Skogan, 2008; Spergel, 2007).

and services.

Two years later, we see divergent effects depending on initial levels of state presence. On average, across the full sample, the intervention did not improve residents' perceptions of state governance and legitimacy. On the contrary, we see some indications that relative state governance *fell* as a result of the intervention.

Since we anticipated that the policy could have different effects depending on the initial conditions, we prespecified a single heterogeneity analysis: by initial levels of state governance. We find that people's opinions of the state fell in neighborhoods with weaker initial state presence (in absolute terms, and also relative to the gang). In places the state was initially stronger, however, residents' opinions of the state improved.

Post-treatment interviews with community leaders, combo members, and municipal liaisons give some clues why. They suggest that, where initial state presence was weak, the central task force was unable to deliver on important promises. Related, the presence of the liaisons may even have raised expectations of what the task force and municipal apparatus could actually deliver, and so failures to follow through were doubly disappointing.

The evidence is not consistent with a handful of other potential channels. For instance, we see no systematic evidence that combos reacted to the intervention, either by trying to co-opt or sabotage the liaisons. Nor do we see indications that combos tried to compete with the city for citizen loyalty by increasing their own service-provision. Rather, it seems that if the state failed to live up to its promises, it did so on its own terms. Related work observes a strategic reaction in a quasi-experimental study of how combos responded to increases in security services over 30 years (Blattman et al., 2022a). Possibly this is a response to policing. Alternatively, in this experiment, had the city government not stumbled in the neighborhoods where they were weakest, then it is also possible that the liaisons would have provoked a strategic criminal response.

These results could shed light on a common feature of cities worldwide: high government attention to places where the state is strong, and persistent neglect of the places where

the state is weak. If short-term returns to effort are lowest in the places with little state capacity (or where gangs govern), then there are both bureaucratic and political reasons to neglect the weaker spaces. The bureaucratic one can arise from state agents taking the path of least resistance—a principal-agent problem within the state. Politicians must also decide where to invest their marginal effort and resources, and may direct their efforts to the areas where visible progress is easiest to achieve. This could help explain the heterogeneity of state presence we see, even in the metropolises of middle-income nations, and even where governments have made efforts to address the problem.

This would be consistent with the idea that there are increasing returns to state capacity over some range. Returns to investments in state capacity could be low (or even negative) at low initial levels of state presence, and increasing into a positive range at middle levels (attracting further state investment to the already rich zones), while ultimately become decreasing beyond some point when state presence is firmly established. Increasing returns and path dependence are a common feature of many other aspects of political and economic development (Pierson, 2000; Romer, 1986; Banerjee and Duflo, 2011), and the concept could apply to statebuilding investments as well. We see this is an important hypothesis for future research.

2 Context

Medellín is Colombia's second-largest metropolitan area, with a population of almost 4 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.

Like most large Colombian cities, Medellín has a well-organized bureaucracy with high tax revenues and public services. This includes a large and professional Metropolitan Police force—a branch of the National Police. Each community also has an elected community action board that helps local groups regulate and organize their community, plus churches and other local organizations. Most public services, however, are provided by the Alcaldía

and its various Secretariats.

One of the largest municipal agencies is the Secretariat of Security—an organization of roughly 2,500 staff who provide numerous services to residents, including responding to various emergencies and street disorder, directly resolving community disputes and domestic violence, and regulating the use of public space. It sits directly beneath the Mayor and is the city’s primary organization for setting security policy and investing in security infrastructure.

In addition to these state and community security organizations, virtually every low- and middle-income neighborhood in Medellín has a local gang called a *combo*. We conducted a census of *combos*, in which we identified nearly 400 in the metropolitan area, with a main street corner or known location for each (see Blattman et al. (2022b) and Appendix A). This experiment emerged from a larger and longer-term qualitative study of *combos* and their organizations. As part of this larger project, over six years we have interviewed 118 criminal leaders and members across 41 groups.³ This section’s discussion of *combo* operations and governance are based on this long-run observation and analysis.

Most *combos* have 15 to 50 permanent, salaried members between the ages of 15 and 35. Their territories (often no more than a few dozen blocks) tend to be long-standing, well-defined, known to locals, and relatively stable over time. Many *combos* have been in their neighborhood for generations, and members come from the local area.

To earn money, *combos* monopolize local illegal markets, especially retail drug sales, prostitution, and local loan-sharking. They frequently participate in and regulate local legal markets in consumer goods, especially cooking gas, arepas, milk, and eggs. They also extort local businesses and sometimes residents, construction sites, and bus companies that operate

³To enable evaluation of these data, we highlight the sampling, methods, and ethical measures taken.

To begin, note that this is a convenience sample of criminal actors. Half the interviews were in prison. We believe our subjects 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.

We had several strategies for maintaining confidentiality of criminal group members. Above all, we were transparent about our research aims and work with the government. We 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.

routes through a combo's territory.

Many residents of Medellín can also turn to combos for everyday forms of governance. Based on several years of qualitative work in these neighborhoods, we identified 17 governance services that both the state and the combos commonly provide to residents and businesses. In 2019 we surveyed nearly 7,000 residents and businesses on the degree to which their neighborhood combo provides these services, as well as the perceived legitimacy of both, and levels of taxation and other payments to combos. The city is divided into roughly 250 areas called *barrios*, and the survey was representative of all 223 low- and middle-income barrios.

To measure governance, we asked residents how frequently each actor responded to these 17 common disputes and forms of disorder (12 from residents and 5 from business-owners). Table 1 reports scaled responses, where 0 = Never, 0.33 = Occasionally, 0.66 = Frequent, 1 = Always. We create average indexes of *State* and *Combo governance* (0 to 1), as well as the difference between them, *Relative state-versus-combo governance*, which can vary from -1 to 1.

The average response for any service by either provider was seldom greater than 0.5, suggesting that neither the state nor the combo are regularly responsive. In relative terms, combo response was generally lower than the state's, but higher in five situations: muggings and theft prevention, business and household debt collection, and street fights.

These averages conceal a great deal of variation across combos and barrios, however. Figure 1 maps relative state governance by barrio. The state is present in every neighborhood, but varies in its responsiveness and penetration. A combo is almost always present, but combos vary widely in the extent to which they offer governance and security services. Many choose to provide no governance at all. Similarly, the state's ability to project power varies across the city. As a result, while the state is the dominant provider of protection in most neighborhoods, we observe wide variation.

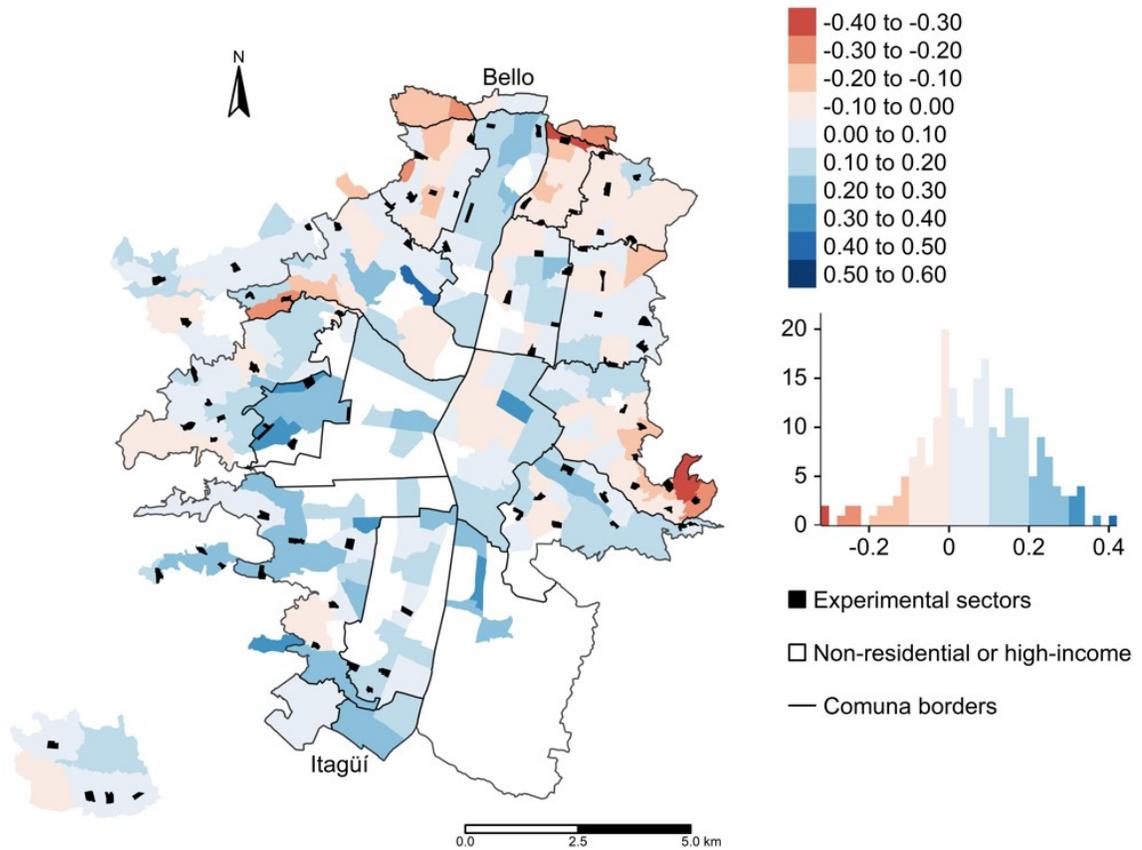
The state offers a different kind of governance, of course. The state's dispute resolution

Table 1: State and combo governance and legitimacy, barrio survey averages, 2019

	Frequency/Rate (0-1 Scale)				Relative State – Combo	
	State		Combo		City-wide survey	Exp. sample
	Estimate	SD	Estimate	SD		
	(1)	(2)	(3)	(4)		
Governance Index	0.41	0.26	0.34	0.29	0.07	0.05
How often they intervene when:						
HH: Someone is making noise	0.43	0.38	0.19	0.30	0.23	0.25
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	0.13
Biz: You have to react to a robbery	0.52	0.37	0.40	0.39	0.11	0.08
Biz: It is necessary to prevent a theft	0.45	0.37	0.38	0.39	0.07	0.04
Biz: Businesses in this sector are robbed	0.42	0.39	0.35	0.38	0.05	0.04
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	-0.00
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	-0.03
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	-0.05
HH: Kids fight on the street	0.29	0.35	0.32	0.37	-0.04	-0.04
Biz: Someone does not want to pay a debt	0.17	0.31	0.23	0.35	-0.06	-0.06
HH: Someone refuses to pay a big debt	0.22	0.31	0.39	0.38	-0.16	-0.21
Legitimacy Index	0.58	0.21	0.43	0.28	0.13	0.14
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 trust the...	0.59	0.23	0.50	0.29	0.09	0.10
How much do your neighbors trust the...	0.57	0.28	0.47	0.36	0.09	0.08

Notes: The governance and legitimacy 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 224 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 governance scales correspond to: 0 = Never, 0.33 = Occasionally, 0.66 = Frequently, 1 = Always. All legitimacy scales correspond to: 0 = Nothing, 0.33 = A little, 0.66 = Somewhat, 1 = Very. Both households (HH) and businesses (Biz) were surveyed on governance levels, but only households were surveyed on legitimacy (hence there are fewer observations).

Figure 1: Experimental sectors and relative state governance by barrio



Notes: The figure displays relative state governance for each low- and middle-income barrio, using the average of all 17 items from Table 1, averaging across all survey respondents in the barrio. (We did not survey high-income barrios.) We also depict the shape of the 80 experimental sectors in black. They are widely spread across the city, but note that there are none in the more central areas, which are more commercial or industrial.

and court systems tend to be impartial and professional, and city leaders are elected in competitive elections every four years. The combo is unelected and relatively unaccountable, and provides “justice” in favor of those who hire them or who are closest to them. At the same time, combos have more local knowledge and deeper networks than most state bureaucrats. Combos are also available all the time, and act swiftly. Thus, 67% of survey respondents said the combo was easy to contact compared to 63% for the police and 32% for the Alcaldía. They also said the combo responded rapidly 58% of the time compared to 41% for the police and 27% for the Alcaldía.

For its services, the state collects fees and taxes from businesses and residents. Likewise, when combos provide services, they typically do so at a price. They fine people who fight or cause disturbances on the street.

Finally, both the state and the combo enjoy a reasonably high level of legitimacy in the eyes of residents. The survey asked residents how much they trust each actor; whether each actor was fair; whether residents were satisfied with each actor; and whether residents thought their neighbors trust and are satisfied with each actor. Table 1 reports scaled responses for the legitimacy questions. As we do with governance questions, we create average indexes of *State* and *Combo legitimacy* (0 to 1), as well as the difference between them, *Relative state-versus-combo legitimacy*, which can vary from -1 to 1 . On average, residents rate their trust and satisfaction of the combo lower than the state, although the difference is not large.

3 Experimental procedures

We worked with the city to expand and experimentally evaluate an existing intervention. In 2017, we found a little-known effort in one of Medellín’s under-served barrios, with a population of roughly 20,000 and a strong gang presence. A small unit in the government sent 7 outreach staff to the barrio. From 2012–17, these “liaisons” set out to build and improve civil society organization and connect residents to existing city services. Based on

our community interviews and observation, it appeared that citizen use of state services increased and that access to and legitimacy of the state rose. Our interviews with combo members in other parts of the city suggested that they would respond to city presence with relief, as they saw governance as a tedious and unprofitable service.

3.1 Sample

The Alcaldía identified 80 “sectors” from its low- and middle-income barrios, choosing ones with varying levels of both state and combo governance. All sectors had a strong gang presence, as did virtually every low- and middle-income neighborhood in the city. A sector is an informal neighborhood, far smaller than a barrio, usually with about 200–600 households with 1,000–3,000 residents, and comprising 5–10 medium-density city blocks. The 80 sectors in our experimental sample are fairly representative of the variation across Medellín’s low- and middle-income barrios in terms of their relative state and combo governance, as seen in column 6 of Table 1 and Appendix Figure B.1. Figure 1 displays the sectors.

We intentionally kept the sample of sectors small in order to achieve the desired level of intensity, as the city’s short-term budget would limit them to hiring 40 new staff. We also wanted to minimize the possibility of spillovers, and growing the number of treated sectors would have raised the risk of contamination.

3.2 Intervention

The city intensified normal municipal services in 40 of the 80 sectors for 20 months, beginning April 2018. Control sectors received normal services.

First, at the city level, the Mayor’s office created an inter-agency task force to respond to local concerns—including poor trash pickup, broken playground equipment, or a lack of attention from the city’s dispute resolution officers. Relatedly, city officials also attended semi-annual formal government-community meetings in the treated sectors, known as *Consejos de Convivencia*, where they and community members would agree on a formal list of commitments. They also organized a large one-time event called *Caravana de la Conviven-*

cia—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 provide some.

Second, the city also assigned a full-time street-level bureaucrat, a liaison, to each treated sector. Their responsibilities included: coordinating the communication of local concerns, community-state meetings, and the other events we describe above; helping community organizations coordinate local collective action (e.g., coordinating garbage spots and dog excrement norms); providing training to community leaders in dispute resolution and related skills; proactively identifying individual and neighborhood problems and referring them to the relevant city agency for assistance; communicating the city and police’s recommended guidelines for dealing with and correctly reporting nuisances, misdemeanors, and crimes; and referring residents with interpersonal conflicts to the comuna’s dispute resolution office.

Liaisons were similar to the city’s normal professional staff: university-educated men and women ages 25–35. They had weekly or monthly quotas for the above activities and were held accountable by their supervisors. The liaisons were not so much directly involved in dispute resolution and service delivery as they were advocates, a source of information, and a source of organizational capacity.

One way to characterize the intensity of this intervention is to note that normally the city has one liaison per comuna—about 1 per 540 blocks. For the intervention, the Secretariat of Security assigned one liaison to each treatment sector—about 1 per 9 blocks. In some sectors, this was the first time the sector had any direct street presence by the city government other than police. While the liaison represents a 60-fold increase in street-level staff, the broader range of city agencies and services—that in practice worked as the back-office support for the liaisons—did not increase their efforts to the same degree. Speculatively, we estimate this to be a tenfold increase in normal municipal government presence.

The aim of the intervention was to increase the visibility, accessibility, and speed of state services and improve trust in and satisfaction with the Alcaldía as a result. This approach was

rooted in the conventional wisdom that unmet demand for contract enforcement and lower transaction costs opens up business opportunities for strongmen and gangs. Importantly, the aim was not to directly challenge gang rule or crowd out their services, but simply to better deliver existing city services.

Also importantly, the intervention did not affect police and criminal justice activities. There are several reasons for this, including: an interest in testing a theoretically more focused intervention; a desire to test non-coercive approaches; the basis in an existing, small-scale approach; and the fact that police and prosecutors are outside the Mayor's chain of command. The combos are far more aware of and concerned with police presence than with municipal bureaucrats.

3.3 Compliance

There was a high level of street presence and visibility of the liaisons for almost two years. The program closely monitored liaisons. They had weekly targets and quotas for neighborhood events and resident referrals, and their activities and task force responses were formally logged and geolocated. From these records we know that they spent 3–6 days or evenings per week in their sector, held frequent community events, and generally met their referral quotas, all within the few blocks they were assigned. The research team monitored and interviewed the liaisons and task force as well, and our general impression was one of autonomous, enthusiastic, hardworking efforts.

The liaisons also reported that combos rarely interfered with their work or attempted to take credit for services delivered. The exceptions mostly affected the first few weeks of the intervention and then the implementation ran smoothly. Two-thirds reported no interference whatsoever. The other third mostly said that the combo was mainly watchful, such as observing public events and meetings from a distance. Another liaison described the combo helping her set up for a major event on one occasion. There were few incidents of preventing liaisons from doing their job. The combo prevented two liaisons from entering into the community for the first few weeks, but once they were able to explain their job and

role, the liaisons were permitted to enter and perform their jobs without interference.

As for the task force and broader city government, the liaisons reported that the majority of their municipal requests were met. But we also saw some evidence that municipal agencies struggled to deliver some aspects of the intervention. On a scale of 0 to 1 (from full compliance to complete failure to deliver) liaisons rated the wider state compliance roughly 0.34, meaning the state “sometimes” failed to deliver on the requested support. We return to these performance failures below.

3.4 Data and outcomes

For baseline data on the sectors, in February 2018 we surveyed three officials per sector for their assessment of: relative governance service provided by the combo and the state; relative street presence of the combo and the state; and their perceptions of local security and drug use. We also have rich, geolocated administrative data including distance to various state and criminal headquarters, crimes committed, and demographics.

Outcome data come from our December 2019 city-wide survey are summarized in Table 1 above. In addition to the representative sample of barrios, we surveyed approximately 30 residents and businesses per experimental sector. Our two primary outcomes are relative state-versus-combo governance and relative state-versus-combo legitimacy. We pre-registered the design and outcomes in April 2018, then again prior to final data collection. In addition, we pre-specified heterogeneity analysis by our baseline measure of relative state governance.⁴

3.5 Empirical strategy

We grouped the 80 sectors into 40 matched pairs using five baseline indices: relative state-combo governance, relative state-combo street presence, insecurity perceptions, administrative crime reports, and distance to public services and infrastructure. We then randomized one in the pair to treatment. This produced the expected degree of balance

⁴See the Journal of Development Economics pre-results registered report for the the final analysis plan (<https://drive.google.com/file/d/1QiEegA-GDdO34-QONMTcxe5bD6MO7nFI/view>), and the social science registry for previous rounds (AEARCTR-0002622).

along baseline covariates (Appendix Table B.1).

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

$$Y_{isb} = \alpha + \beta T_s + \Theta X_{isb} + \epsilon_{isb} \quad (1)$$

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; and X is our pre-specified vector of controls, including sector-pair fixed effects (the randomization strata) and the five baseline indices we used to match blocks.⁵ We cluster standard errors at the sector level, the unit of intervention.

With this design, we estimated we were powered to detect improvements in state governance and legitimacy of about 12% with a two-tailed test.

Minimizing the risk of spillovers To reduce the chance of interference between units, we selected sectors at least 250 meters distant from one another. A total of 40 intervention sectors also ensured that increased service delivery would not reduce services in control sectors. Ex-post, we can use our representative city-wide survey to estimate and control for spillover effects, by comparing blocks close to treatment sectors to those close to control sectors.⁶ We see no evidence of systematic spillovers to neighboring areas (Appendix Table B.2).

Addressing measurement error Naturally, we are concerned that citizens may misreport gang activities. They may feel uncomfortable talking to outsiders or embarrassed to admit the role of the combo. Such measurement error could attenuate estimated treatment effects somewhat.

Combos are a routine part of everyday life, however. We refined survey questions after dozens of qualitative interviews, fine-tuning language, questions, and approach to elicit truthful answers. We conducted all interviews anonymously and in private, typically indoors.

⁵We report results for alternative specifications in Appendix Table B.3, discussed below.

⁶Blattman et al. (2021) follow a similar approach to estimate spatial spillover effects resulting from a hot spots policing intervention in Bogotá.

In the context of a secret interview, we believe most respondents answered questions freely and truthfully. Three analyses are consistent with this conclusion.

First, 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 more successful in eliciting honest responses.

Second, we used a survey experiment to assess under-reporting in security fee payment. We asked some respondents directly whether they paid; others we used a randomized-response 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 no differences in payment rates between the approaches, suggesting people did not misreport this topic (see Appendix C).

Third, we found that people who appeared not to want to talk about gang rule or security fees often said “I don’t know” or pass on answering that question. Just 7% of the sample answered in that fashion. If this were driven by worries about the combo, we might expect a correlation between combo governance and the proportion of questions unanswered. We see no such relationship (see Appendix C).

4 Results

4.1 Average treatment effects

Despite the length and intensity of additional state presence, on average we see no evidence the intervention improved opinions of the state in treated areas. Table 2 reports program impacts in the 80 sectors. Unexpectedly, we estimate a 0.024 *decrease* in relative state governance, statistically significant at the 10% level.⁷

Looking at the components of the relative governance index—state and combo governance—we draw similar conclusions from this exploratory analysis. State governance, for example, drops by 0.013—a 3.1% decrease relative to the control group mean, not statistically significant. The confidence interval rules out increases in state governance larger than 1.2% of the control mean. The confidence interval also rules out decreases in combo governance larger than 4.3% of the control mean.

Turning to legitimacy, the impact on relative state legitimacy is in the expected direction but small and imprecise. We estimate a 0.017 increase in relative state legitimacy, not statistically significant. Looking at the components of the relative legitimacy index, the confidence interval rules out increases in absolute state legitimacy greater than 4.2% of the control mean. Similarly, the confidence interval rules out decreases in absolute combo legitimacy greater than 7.7% of the control mean.

We see similar patterns across most of the measures that comprise these indexes (not shown). We also see no evidence of an average treatment effect on other survey measures of state efficacy, such as the speed of response and ease of accessing services (Appendix Table B.4).

4.2 Heterogeneity by baseline state governance

As already noted, in some of the experimental sectors the Alcaldía was already active

⁷If we explore alternative (but not prespecified) control vectors, we still see no evidence of significant improvements in perceptions of relative state governance (Appendix Table B.3). In these robustness checks we either drop all covariates or use the double-lasso method proposed by Urminsky et al. (2019).

Table 2: Program impacts on primary outcomes

	Control Mean (1)	ATE (2)	SE (3)	P-value (4)	N (5)
Relative State Governance Index	0.066	-0.024*	0.014	0.086	2,314
State Governance Index (0-1)	0.413	-0.013	0.009	0.144	2,362
Combo Governance Index (0-1)	0.345	0.009	0.012	0.452	2,316
Relative State Legitimacy Index	0.131	0.017	0.018	0.353	1,845
State Legitimacy Index (0-1)	0.572	0.010	0.007	0.150	1,906
Combo Legitimacy Index (0-1)	0.437	-0.006	0.014	0.690	1,845

Notes: Each row is a separate intent-to-treat estimate of program impacts, estimated via Equation 1. We regress each dependent variable on an indicator for treatment and our pre-specified control vector: 5 baseline variables and sector-pair fixed effects. The unit of observation is the individual survey respondent, and we cluster standard errors at the sector level (the unit of randomization). Both households and businesses were surveyed on governance levels, but only households were surveyed on legitimacy (and hence there are fewer observations).

and visible on the street. In others, residents remarked that this was the first time they had seen a representative of the government in their sector who was not a policeman. Certainly there had never been meetings with officials or a municipal services fair held there before, let alone meetings focused on sector-specific problems. This is why we prespecified that we would examine treatment effects in higher and lower state presence sectors separately—the sole heterogeneity analysis we prespecified.⁸

Note that we do not have representative city-wide survey data from residents at baseline, and so our prespecified measure of initial state strength comes from baseline interviews with three community and city leaders per sector, and their assessment of the *relative* role of the state and the combo in providing everyday governance.⁹

Based on post-treatment feedback, we also constructed a proxy for *absolute* baseline state governance. We do so using a weighted average of distance to state headquarters and other

⁸We specified that we would look at above- and below-median levels of baseline relative state governance. We also committed to report treatment effects in the four major quartiles, although with just 80 sectors that analysis is more exploratory as it is under-powered with only 20 sectors in each subgroup.

⁹Different than our endline measure, at baseline we did not ask independently about service delivery and legitimacy for each actor. Rather, we asked who would residents reach out to in addressing their needs.

Table 3: Program impacts in prespecified subgroups: High vs. low initial state governance

	(A) Subgroups by: <i>Relative</i> baseline governance		(B) Subgroups by: <i>Absolute</i> baseline governance	
	Dependent Variable: Relative governance (1)	Dependent Variable: Relative Legitimacy (2)	Dependent Variable: Relative governance (3)	Dependent Variable: Relative Legitimacy (4)
Program impacts:				
Low baseline governance	-0.033* (0.018)	-0.021 (0.029)	-0.060*** (0.017)	-0.026 (0.022)
High baseline governance	-0.016 (0.020)	0.053** (0.023)	0.021 (0.024)	0.073** (0.031)
Difference	0.017 (0.026)	0.074** (0.037)	0.081*** (0.030)	0.099** (0.038)

Notes: The 80 sectors in the experimental sample were blocked into pairs, and one of each pair was randomly assigned to treatment. Each pair was classified into above or below the median of state governance for two different partitions: relative state governance (Panel A) or absolute state governance (Panel B). Here we report estimated program impacts on each subgroup and the difference between the two group, from Equation 2. The unit of observation is the individual survey respondent, and we cluster standard errors at the sector level (the unit of randomization).

baseline administrative and survey data.¹⁰

We estimate average treatment effects on the subgroups via the simple OLS regression:

$$Y_{isb} = \alpha + \beta T_s + \delta(T_s \times Low_s) + \lambda Low_s + \Theta X_{isb} + \epsilon_{isb} \quad (2)$$

where all variables are the same as in Equation 1, and *Low* is an indicator for neighborhoods with below-median state governance. In that case, β estimates the program impact on places with initially higher state presence, δ estimates the difference between high and low neighborhoods, and the estimated program impact on initially lower state presence neighborhoods is the sum of β and δ .

Table 3 reports the heterogeneous treatment effects. Panel A presents results for sectors above and below median relative state governance. Panel B presents results for sectors above

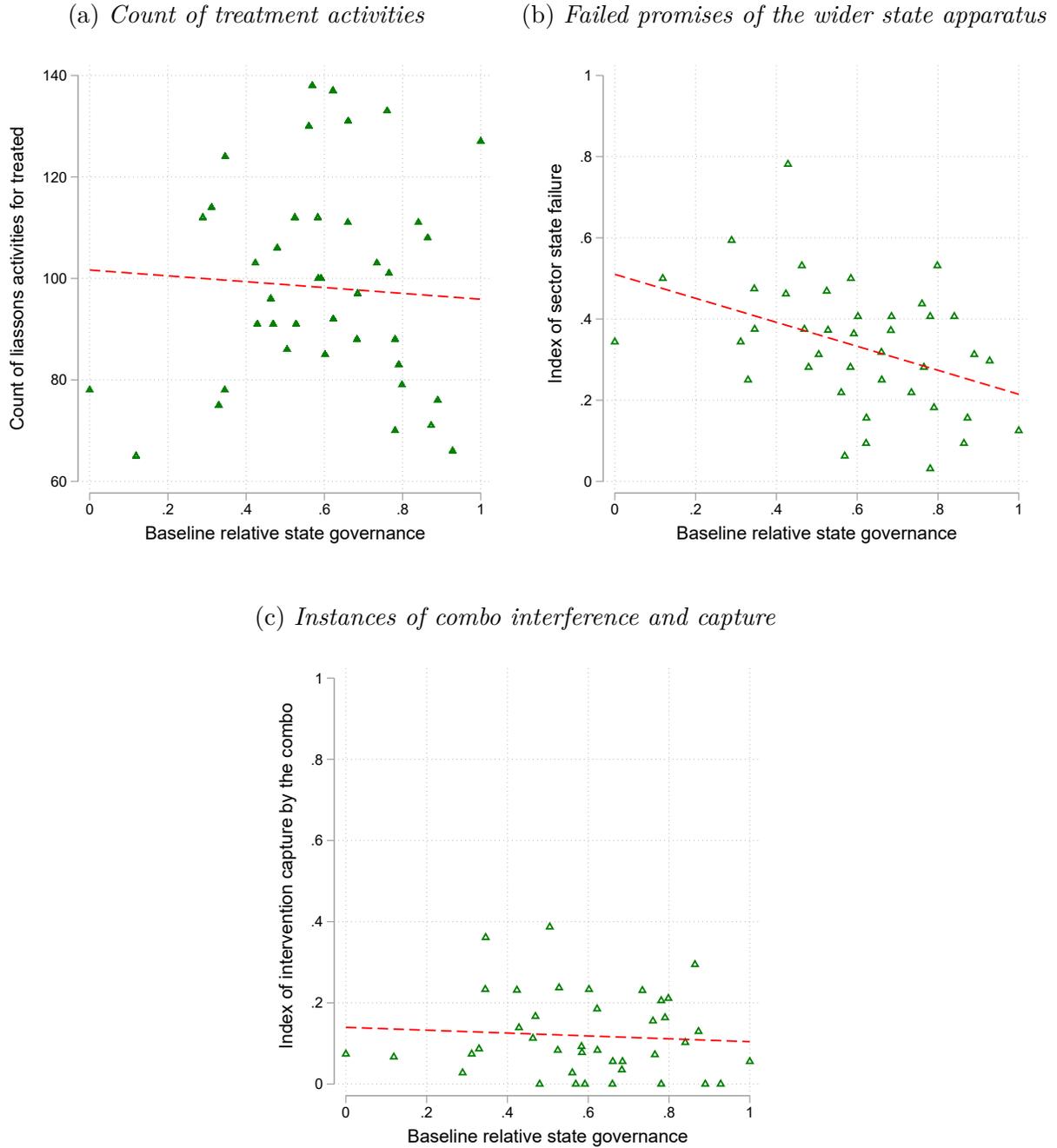
and below median absolute state governance.

We see divergent effects depending on the initial levels of state presence. In the sectors where the state was relatively weak and inactive, treatment is associated with decreases in relative state governance and legitimacy. When we use the baseline measure of relative state governance, we find that relative state governance decreases by 0.03, significant at the 10% level, and relative state legitimacy decreases by 0.02, not significant. In those places where state penetration was relatively high, estimated treatment effects for relative state governance decrease by 0.02, not significant, and relative state legitimacy increases by 0.05, significant at the 5% level. The difference in treatment effects between both groups of sectors is statistically significant at the 5% level for relative state legitimacy.

Qualitatively we see the same subgroup difference when using the measure of absolute state governance to conduct the heterogeneity analysis. In sectors with lower initial state presence the estimated treatment effects imply a decrease of 0.06 in relative state governance (significant at the 1% level) and a decrease of around 0.03 in relative state legitimacy (statistically not significant). In sectors with higher state presence, our estimated treatment effects are an increase of 0.02 in relative state governance (statistically not significant) and an increase by 0.07 in relative state legitimacy (significant at the 5% level). Crucially, the difference in treatment effects in state governance and state legitimacy between the subgroups are statistically generally statistically significant.

These results are robust to different specifications. Appendix Table B.5 shows heterogeneous treatment effects using baseline quartiles of relative state governance, rather than above-below median. We see a similar pattern, with adverse effects concentrating on sectors with less state attention. This table also presents the equivalent heterogeneity results based

Figure 2: How treatment experiences varied by initial levels of gang rule
(treated sectors only)



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. This aggregates several measures: a scale for the frequency and difficulties of interaction with local gangs; a set of binary variables on whether local actors (including the gang) 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 locals gang on intervention activities.

on our baseline measure of absolute state governance, which also suggest a similar pattern.

4.3 Mechanisms

In brief, data from program implementation and qualitative interviews suggest that the street-level staff did their jobs, but in the neighborhoods where the central state had the least historical presence, the central task force and other more central actors had trouble delivering on the promises. Expectations were set, but never met—an insight that we argue might be a general challenge facing state-builders. Finally, we don't see evidence that gang presence or gang governance was the issue facing the state. Combos may have played an unobserved role, but the state may have failed on its own.

First, the administrative data suggest that liaison compliance and effort were high in all neighborhoods. Moreover, initial state governance had little effect on the efforts and activities of liaisons. Panel (a) of Figure 2 uses program data on all events and activities logged by the liaison. It plots the number of documented activities by baseline relative state governance. Activities are numerous and unrelated to initial government presence.¹¹

The same is not true of the Alcaldía's wider activities, including attention from higher-level politicians, bureaucrats, and the task force. Panel (b) of Figure 2 reports the frequency that the central administration failed to deliver on promises. These come from a post-program survey of liaisons. They reported central municipal government failures twice as often in the sectors with relatively low initial state presence. Most common were failures of the city to respond to community needs. Equipment might go unrepaired, for example. Or, as one liaison explained, they organized a meeting between the community and city officials, and the officials never arrived. Another said how they had publicized the new police code—which includes official guidelines for when citizens should call the police versus one of the civilian security and services agencies—but the residents were frustrated because the police

¹⁰We use machine learning methods to identify the baseline covariates more prognostic of endline absolute state governance. Then, we use the coefficients to predict an index and coarsen this measure into an above/below median measure of absolute state governance.

¹¹In a small number of cases, the combo barred the liaison from entering the neighborhood for a period of time. These issues were typically resolved in a handful of weeks and did not affect the larger intervention.

did not follow it reliably.

Taken together, our post-experimental analysis with the liaisons broadly points to a mix of unmet expectations as a plausible mechanism. A related hypothesis is that the presence of liaisons also raised expectations of service delivery, which were hardest to meet in lower state presence areas. But the data suggest that even if expectations were not raised too high, the state apparatus was simply unable to follow through on promises in the neighborhoods where historically they did not have a strong presence, even in absolute terms.

Finally, we see no evidence that combo responses shaped the heterogeneous results. Granted, combos noticed an increased presence of the Alcaldía. In qualitative interviews, for instance, almost all liaisons described having to explain their presence to the combo. Thus, we are confident that combos were generally aware of increased state activity from the beginning.

Most of the evidence, however, suggests that the combos did not react to the presence of these liaisons and the attention of the task force. For example, Panel (c) of Figure 2, captures the degree with which liaisons reported that the combo interfered with their activities.¹² We do not see much evidence that the combo tried to capture the liaison's activities or take credit for their work. The levels are low and there is little relationship with initial state presence. Nor do we see any evidence that combos escalated their governance services or legitimacy in response to the state. The coefficients on the combo indexes in Tables 2 and 3 are generally close to zero. We do not want to understate the role of combos, since low initial state presence could be endogenous to gang strength. But there is nothing in our results to

¹²See the foot of the figure for details on how we built the intervention capture measure.

suggest direct interference.¹³

4.4 Addressing measurement error concerns

Finally, could measurement error account for these results? The evidence suggests not. 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. To explain our results, residents in treated sectors would have needed to systematically under-report state governance or overstate combo governance, and only in the treatment sectors that had low initial government presence—in essence, the reverse of normal experimenter demand. We test for any correlation between treatment and measurement error in Appendix Table C.1. We have two proxies for measurement error—non-response to combo-related questions relative to responses about the state, and the randomized response survey experiment. We do not see significant evidence of misreporting correlated with treatment.

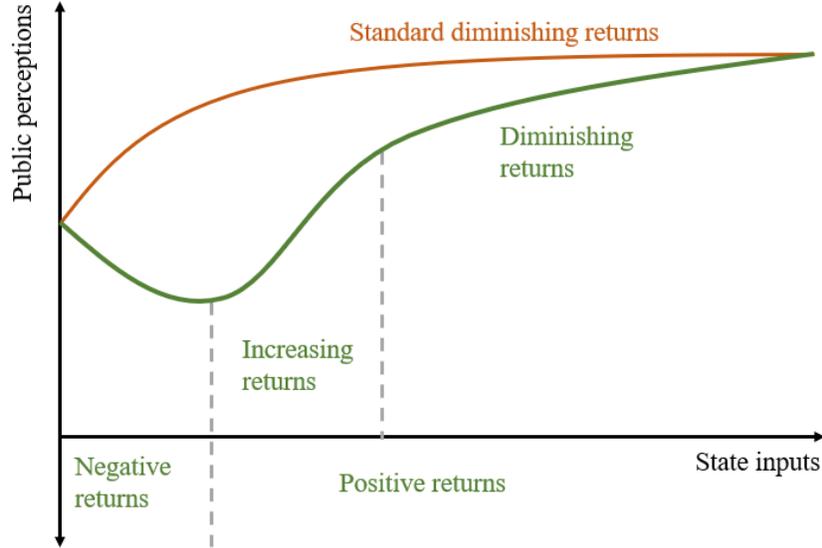
5 Discussion and conclusions

Nearly two thirds of the world lives in cities. Municipal governments typically provide the most basic and everyday public services—lights, water, property rights, and security and public order. Many neighborhoods receive few government services at all. A natural question is “why?” In principle, elected officials ought to have incentives to offer voters these services.

The results of this experiment in Medellín suggest an intriguing hypothesis for the variation in state penetration: there may be nonlinear returns to investments in state-building,

¹³This is notable partly because, in a companion study of long term combo reactions to increased policing, Blattman et al. (2022a) interviewed gang leaders and learned that some see themselves as directly competing with the state for civilian loyalty. Gangs value this loyalty as it protects their drug rents. Using a simple model of imperfect competition in the market for security and protection, that paper shows that a combo solely concerned with selling protection will reduce its governance services in response to an increase in competition from the state. Once you introduce the drug market, however, and the returns to loyalty and legitimacy, the optimal response of the combo may switch, and it will elevate its governance services in response to state intervention. That paper quasi-experimentally studies a large, sustained increase in policing and related security services over 30 years, and sees evidence that combo’s respond strategically to state presence by governing more.

Figure 3: Stylized depiction of diminishing and non-linear returns to state investments



depending on the starting point. In particular, marginal returns to some of the initial investments may be low or negative, to the extent that local experience, presence, structures and accountability are important for effective service delivery.

Our initial hypothesis was that there were positive but diminishing returns to state effort and inputs. But the findings above suggest that, in places with low initial levels of state investment, the marginal returns to investment were negligible or even negative, in part because the state had trouble following through. We illustrate the distinction in Figure 3 with two highly-stylized “production functions”, $F(I)$, for how state inputs I shape perceptions of state governance and legitimacy. The upper curve depicts a standard diminishing-returns function, consistent with our prior beliefs and the conventional wisdom, with $F'(I) > 0$ and $F''(I) < 0$ everywhere. The lower S-shaped curve, however, is more consistent with our results, and suggests zones of negative and positive marginal returns (i.e. initially $F' < 0$, before switching to $F' > 0$), as well as increasing returns over a certain range ($F'' > 0$) before diminishing returns ($F'' < 0$) set in, all depending on the initial level of state investment.

A related possibility is that the coercive arm of the state needs to increase as well—meaning that municipal services alone are not enough, or are not the main source of state

legitimacy (especially when gangs are strong). Thus, the S-curve could arise because at low levels of state investment there are complementarities between municipal services and policing and justice systems. This would be consistent with a literature on counter-insurgency, which has argued that a combination of military action followed by state service provision increases state legitimacy and civilian collaboration against the insurgents (Albertus and Kaplan, 2013; Berman et al., 2011, 2013; Berman and Matanock, 2015). On the other hand, increased policing could provoke a strategic response from competing armed actors, as with combos in Blattman et al. (2022a), attenuating the effects on state relative governance and legitimacy.

Altogether these hypotheses suggest that in the places a state is weakest, building capacity and legitimacy may require longer, larger, and more multifaceted interventions than previously thought. This is a hypothesis, but an important one for future research, because it could help explain why so many cities have patchwork state presence. Baseline state strength is seldom uniform, so if broad-brush efforts are likely to lead to larger improvements in places with better baseline conditions then initial inequities are likely to grow starker. Another implication is that short-term electoral incentives may run against investing in state capacity in the neighborhoods with the weakest state presence. Governments and international organizations may need to think more about how to design institutions, term limits, federal aid, and international assistance to incentivize long run investments in state capacity and reducing the influence and legitimacy of other armed actors in society.

References

- Acemoglu, D., A. Cheema, A. I. Khwaja, and J. A. Robinson (2020). Trust in state and nonstate actors: Evidence from dispute resolution in pakistan. *Journal of Political Economy* 128(8), 3090–3147.
- Acemoglu, D., G. De Feo, and G. D. De Luca (2020). Weak States: Causes and Consequences of the Sicilian Mafia. *The Review of Economic Studies* 87(2), 537–581.

- Albertus, M. and O. Kaplan (2013). Land Reform as a Counterinsurgency Policy: Evidence From Colombia. *Journal of Conflict Resolution* 57(2), 198–231.
- Arias, E. D. (2006). The Dynamics of Criminal Governance: Networks and Social Order in Rio de Janeiro. *Journal of Latin American Studies* 38(2), 293–325.
- Arjona, A. (2016). *Rebelocracy*. Cambridge University Press.
- Banerjee, A. and E. Duflo (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. Public Affairs.
- Berman, E., J. H. Felter, J. N. Shapiro, and E. Troland (2013). Modest, Secure, and Informed: Successful Development in Conflict Zones. *American Economic Review* 103(3), 512–17.
- Berman, E. and D. D. Laitin (2008). Religion, Terrorism and Public Goods: Testing the Club Model. *Journal of Public Economics* 92(10-11), 1942–1967.
- Berman, E. and A. M. Matanock (2015). The Empiricists’ Insurgency. *Annual Review of Political Science* 18, 443–464.
- Berman, E., J. N. Shapiro, and J. H. Felter (2011). Can Hearts and Minds be Bought? The Economics of Counterinsurgency in Iraq. *Journal of Political Economy* 119(4), 766–819.
- Blattman, C., G. Duncan, B. Lessing, and S. Tobón (2022a). Gang Rule: Understanding and Countering Criminal Governance. *Working Paper*. https://www.nber.org/system/files/working_papers/w28458/w28458.pdf.
- Blattman, C., G. Duncan, B. Lessing, and S. Tobón (2022b). Gangs of Medellín: How Organized Crime is Organized. *Working Paper*.
- Blattman, C., D. Green, D. Ortega, and S. 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, C., A. C. Hartman, and R. 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.

- Cammett, M. and L. M. MacLean (2014). *The Politics of Non-State Social Welfare*. Cornell University Press.
- Craig, W. M., F. Vitaro, L. Gagnon, and R. E. Tremblay (2002). The Road to Gang Membership: Characteristics of Male Gang and Nongang Members from ages 10 to 14. *Social Development* 11(1), 53–68.
- Cureton, S. (1999). Gang Membership: Gang Formations and Gang Joining. *Journal of Gang Research* 7(1), 13–21.
- Curry, G. D., S. H. Decker, and A. Egley Jr (2002). Gang Involvement and Delinquency in a Middle School Population. *Justice Quarterly* 19(2), 275–292.
- Decker, S. H. and G. D. Curry (2000). Addressing Key Features of Gang Membership: Measuring the Involvement of Young Members. *Journal of criminal justice* 28(6), 473–482.
- Del Castillo, G. (2008). *Rebuilding War-Torn States: The Challenge of Post-Conflict Economic Reconstruction*. OUP Oxford.
- Esbensen, F.-A., D. W. Osgood, T. J. Taylor, D. Peterson, and A. Freng (2001). How great is great? results from a longitudinal quasi-experimental design. *Criminology & Public Policy* 1(1), 87–118.
- Gambetta, D. (1996). *The Sicilian Mafia: the Business of Private Protection*. Harvard University Press.
- Ghani, A. and C. Lockhart (2009). *Fixing Failed States: A Framework for Rebuilding a Fractured World*. Oxford University Press.
- Gray, O. (2003). Badness-Honour. *Understanding Crime in Jamaica: New challenges for Public Policy*, 13–48.
- Henn, S. J. (2021). Complements or Substitutes? How Institutional Arrangements Bind Chiefs and the State in Africa.
- Higginson, A., K. Benier, Y. Shenderovich, L. Bedford, L. Mazerolle, and J. Murray (2015). Preventive interventions to reduce youth involvement in gangs and gang crime in low-and

- middle-income countries. *Campbell Systematic Reviews* 11(1), 1–176.
- Karim, S. (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.
- Kasfir, N. (2015). *Rebel Governance—Constructing a Field of Inquiry: Definitions, Scope, Patterns, Order, Causes*, Volume 2. Cambridge University Press New York.
- Lessing, B. (2020). Conceptualizing Criminal Governance. *Perspectives on Politics*, 1–20.
- Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review* 94(2), 251–267.
- Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy* 94(5), 1002–1037.
- Sánchez De La Sierra, R. (2020). On the Origins of the State: Stationary Bandits and Taxation in Eastern Congo. *Journal of Political Economy* 128(1), 000–000.
- Skaperdas, S. (2001). The Political Economy of Organized Crime: Providing Protection When the State Does Not. *Economics of Governance* 2(3), 173–202.
- Skarbek, D. (2011). Governance and Prison Gangs. *American Political Science Review* 105(4), 702–716.
- Skogan, W. G. (2008). Brief Summary: An Evaluation of CeaseFire-Chicago. *Working Paper*. http://www.skogan.org/files/CeaseFire_Brief_Summary.pdf.
- Spiegel, I. A. (2007). *Reducing Youth Gang Violence: The Little Village Gang Project in Chicago*. Rowman Altamira.
- Tilly, C. (1990). *Coercion, Capital, and European States, AD 990-1990*. Blackwell.
- Urminsky, O., C. Hansen, and V. Chernozhukov (2019). The double-lasso method for principled variable selection. *Working Paper*. [10.31234/osf.io/2pema](https://doi.org/10.31234/osf.io/2pema).
- Van der Windt, P., M. Humphreys, L. Medina, J. F. Timmons, and M. Voors (2019). Citizen Attitudes Toward Traditional and State Authorities: Substitutes or Complements? *Comparative Political Studies* 52(12), 1810–1840.

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

Weigel, J. L. (2020). The Participation Dividend of Taxation: How Citizens in Congo Engage More with the State When it Tries to Tax Them. *The Quarterly Journal of Economics* 135(4), 1849–1903.

Appendix

Table of contents

A Combo census	ii
B Supplemental analysis	iv
C Assessing measurement error in outcomes	viii

A Combo census

The authorities have never developed a systematic list of combos. To create our census, we began by integrating five existing lists that are known to be incomplete:

1. A public list of 246 combos from the leading local newspaper *El Colombiano* from 2014, updated in 2017;
2. A confidential list of 96 combos from the Medellín's Mayor's Office from 2010;
3. A confidential list of 190 combos from the Office of the Attorney General from 2015;
4. A confidential list of 87 combos from the Medellín's Mayor's Office from 2017; and
5. A confidential list of 89 combos from the Medellín's Mayor's Office, the Office of the General Attorney, and the National Police from 2018.

We then tried to validate the existence of the combos as well as identify a prime location, usually a drug selling spot or their headquarters. We did so using primary contacts with current and former combo and razón members, local experts (especially journalists and community leaders), as well as with law enforcement officials. The principal challenges here included:

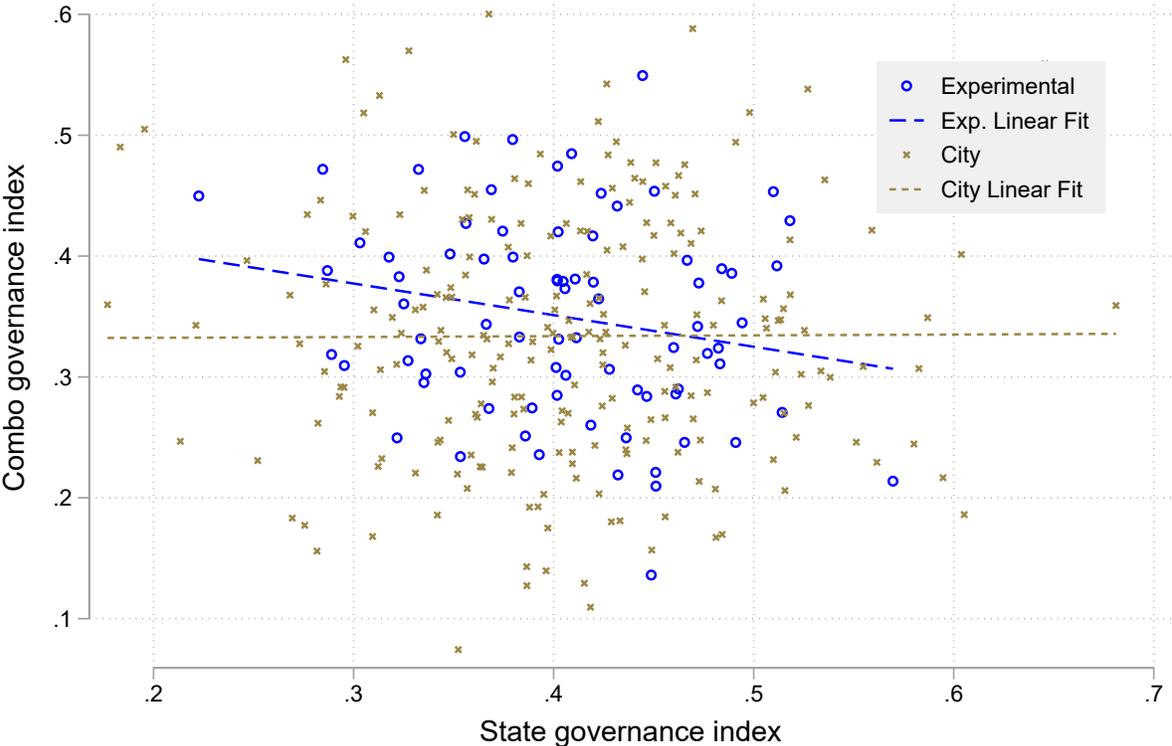
- distinguishing between razones and combos;
- determining what groups were actually "sub-combos"—departments or subdivisions within a single combo;
- identifying cases of multiple names for the same combo (these may be historical names, nicknames, or police/journalist error in one of the lists; and
- assessing defunct combos that have disbanded or absorbed into another combo.

Altogether, this exercise has some limitations, including:

- We did not physically search the streets for combos, or speak directly to residents and leaders in a systematic way, and confirm combos in that matter. Our many qualitative interviews with residents, leaders, combo members and experts did inform our classification, however.
- We also did not systematically count and verify combos in neighboring municipalities, and so 50 combos in the rest of the metropolitan area is a rough estimate.
- Our merger and validation of the existing lists relies heavily on the expertise and judgment of 3 professional experts, journalists, and former criminal group members.

B Supplemental analysis

Figure B.1: Endline governance levels in the representative city sample and the experimental sample



Notes: The figure plots average 2019 state and combo governance levels in each city barrio as well as the 80 experimental sectors. 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.

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

Covariate	Means		Regression Difference		
	Control	Treated	Coeff	p-value	SE
<i>Baseline indices used for matching , main control vector</i>					
Additive index of combo presence and governance	0.00	-0.02	-0.02	0.92	0.22
Baseline - Combo Governance Index (Relative to State)	0.00	-0.02	-0.02	0.91	0.22
Standardized index of perceived insecurity and drugs	0.06	-0.07	-0.13	0.58	0.22
Index of crime	0.09	-0.12	-0.21	0.35	0.22
Index of distance from public goods and services	-0.14	0.14	0.28	0.22	0.22
<i>Other baseline covariates</i>					
Respondent age between 18 and 25	0.19	0.19	-0.00	0.98	0.02
Respondent age between 26 and 40	0.29	0.31	0.01	0.52	0.02
Respondent age between 41 and 64	0.39	0.37	-0.01	0.58	0.02
Respondent is business owner	0.20	0.20	0.00	0.90	0.00
Multidimensional Poverty Index (2018)	14.34	17.26	2.93	0.20	2.26
Block Longitude	-75.58	-75.58	-0.01	0.37	0.01
Block present in 1970	0.50	0.44	-0.06	0.54	0.10
Median age (2005)	27.20	26.31	-0.90	0.41	1.08
Total women (2005)	133.86	142.04	8.18	0.53	12.90
Total non-mestizo population (1993)	0.53	0.61	0.08	0.66	0.18
Median age (1993)	24.16	24.71	0.56	0.60	1.06
Share of women (1993)	0.53	0.52	-0.01	0.68	0.01
Distance to the respective razon headquarters (100 meters)	17.28	19.12	1.84	0.66	4.15

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 B.2: Estimating treatment spillovers onto blocks within a 250 meter radius

	Treatment Estimate	P-value	0m-250m Spillover Estimate	P-value
	(1)	(2)	(3)	(4)
Relative State Governance Index	-0.0309	0.1210	-0.0672	0.9190
State Governance Index (0-1)	-0.0149	0.2320	-0.0299	0.9460
Combo Governance Index (0-1)	0.0140	0.3780	0.0312	0.8690
Relative State Legitimacy Index	0.0057	0.8890	-0.0511	0.7060
State Legitimacy Index (0-1)	0.0107	0.3410	-0.0108	0.5830
Combo Legitimacy Index (0-1)	0.0061	0.7760	0.0362	0.8470

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 table 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. As Blattman et al. (2021) note, 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.

Table B.3: Robustness of experimental results to changes in the control vector

	Specification of control vector		
	Main (5 baseline matching controls + sector-pair FE)	No controls	Lasso-selected controls No FE
	Estimate (SE)	Estimate (SE)	Estimate (SE)
	(1)	(2)	(3)
Governance variables			
Relative State Governance Index	-0.024* (0.014)	0.002 (0.011)	-0.001 (0.016)
State Governance Index	-0.013 (0.009)	0.009 (0.010)	0.005 (0.012)
Combo Governance Index	0.009 (0.012)	0.005 (0.012)	0.005 (0.015)
Legitimacy variables			
Relative State Legitimacy Index	0.017 (0.019)	0.016 (0.022)	0.022 (0.025)
State Legitimacy Index	0.010 (0.007)	0.010 (0.010)	0.016 (0.012)
Combo Legitimacy Index	-0.006 (0.014)	-0.013 (0.015)	-0.011 (0.018)

Notes: Column (1) replicates our main specification, as reported in Table 2, which includes 5 baseline covariates, as well as 40 sector-pair fixed effects. Column (2) reports treatment effects without any covariates. Column (3) uses the double-lasso method of covariate selection, but excludes from the choice set the 40 sector-pair fixed effects.

Table B.4: Impacts of treatment on survey measures of state efficacy

	Control Mean	ATE	SE	P-value	N
	(1)	(2)	(3)	(4)	(5)
State efficacy index	0.505	-0.008	0.008	0.292	1,907
How easy is to contact the combo	0.453	-0.004	0.009	0.627	1,881
How would this sector be without the combo	0.683	-0.006	0.008	0.457	1,880
How fast is the State	0.381	-0.013	0.011	0.205	1,879
Combo efficacy index	0.547	-0.004	0.012	0.758	1,790
How easy is it to contact the combo	0.591	0.011	0.016	0.499	1,649
How would this sector be without the combo	0.524	-0.014	0.014	0.337	1,706
How fast is the combo	0.560	0.002	0.016	0.887	1,589

Notes: This table calculates the treatment effects on 6 measures of efficacy, and indexes constructed for these measures, using the same approach as in Table 2.

Table B.5: Subgroup analysis: Quartile analysis

	(A) Subgroups by: <i>Relative</i> baseline governance		(B) Subgroups by: <i>Absolute</i> baseline governance	
	Dependent Variable: Relative governance (1)	Dependent Variable: Relative Legitimacy (2)	Dependent Variable: Relative governance (3)	Dependent Variable: Relative Legitimacy (4)
Program impacts:				
Q1 (0 th - 25 th quartile baseline rel. gov)	-0.025 (0.031)	-0.072* (0.038)	-0.058** (0.024)	-0.008 (0.037)
Q2 (25 th - 50 th quartile baseline rel. gov)	-0.039** (0.018)	0.020 (0.037)	-0.063*** (0.022)	-0.044* (0.026)
Q3 (50 th - 75 th quartile baseline rel. gov)	0.026 (0.026)	0.051 (0.033)	0.087*** (0.020)	0.127*** (0.046)
Q4 (75 th - 100 th quartile baseline rel. gov)	-0.067* (0.034)	0.052 (0.040)	-0.047 (0.037)	0.015 (0.041)
Differences relative to Q1:				
Q2	-0.014 (0.035)	0.093* (0.049)	-0.006 (0.033)	-0.036 (0.046)
Q3	0.051 (0.040)	0.123** (0.048)	0.145*** (0.030)	0.134** (0.056)
Q4	-0.042 (0.046)	0.124** (0.058)	0.011 (0.046)	0.023 (0.058)

Notes: This table replicates the results of Table 3 but partitioning the sample in 4 subgroups (quartiles) as opposed to 2. Columns 1 and 2 replicate Panel A of Table 3, and columns 3 and 4 replicate Panel B. 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. We regress each dependent variable on an indicator for treatment and our pre-specified control vector: 5 baseline variables and sector-pair fixed effects. Additionally, we have an indicator variable for each of the three higher quartiles and a variable interacting those indicators with treatment. The unit of observation is the individual survey respondent, and we cluster standard errors at the sector level (the unit of randomization). Both households and businesses were surveyed on governance levels, but only households were surveyed on legitimacy (and hence there are fewer observations).

C Assessing measurement error in outcomes

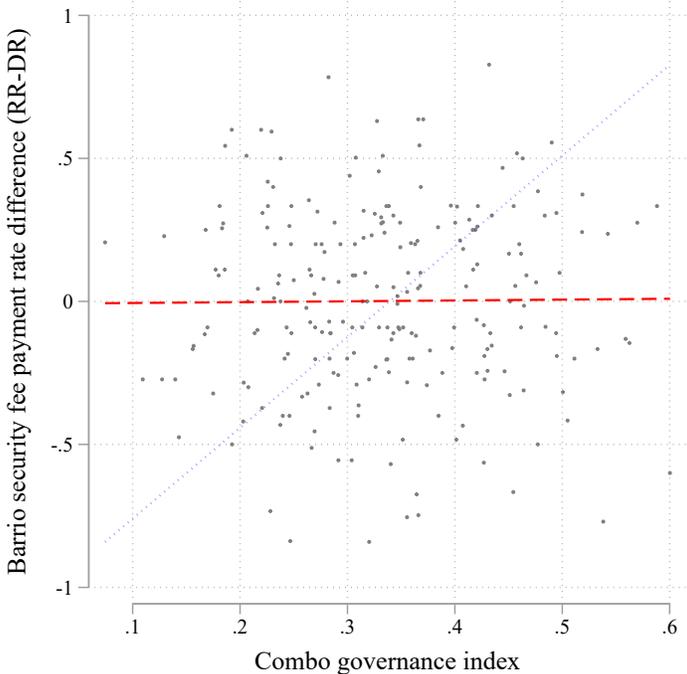
Vacuna survey experiment Paying vacunas seemed to be one of the more sensitive questions on the survey, according to our qualitative experiences. To test this, we randomized how we asked the question. Some respondents were asked directly whether they paid a regular vacuna (Direct Response, or DR); others were asked to use a Randomized Response (RR) technique, where they privately flipped a coin and reported honestly only if it is heads.

We see little statistically significant differences across the two methods. 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 formally examine whether there are treatment-control differences in Table C.1, and we see no indication of such differences.

In Figure C.1 we calculate 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.

Patterns of non-response We also examine patterns of non-response. One concern we had in piloting the survey was that respondents who do not want to talk about the combo may say “I don’t know” or pass on answering that question, and enumerators are permitted to skip questions. Just 7% of the sample answered don’t know or skipped at least 25% of the combo services questions. If this were primarily driven by worries about combo, we might expect a correlation between combo governance and the proportion of questions unanswered. Figure C.2 plots each respondent’s percentage of answered questions against barrio-level measures of combo and mayoral governance levels. We see no substantively or statistically significant correlations. Control group members answer 85 to 97% of sensitive questions regarding the combo. This is 0.2 to 1.1 percentage points lower in the treatment

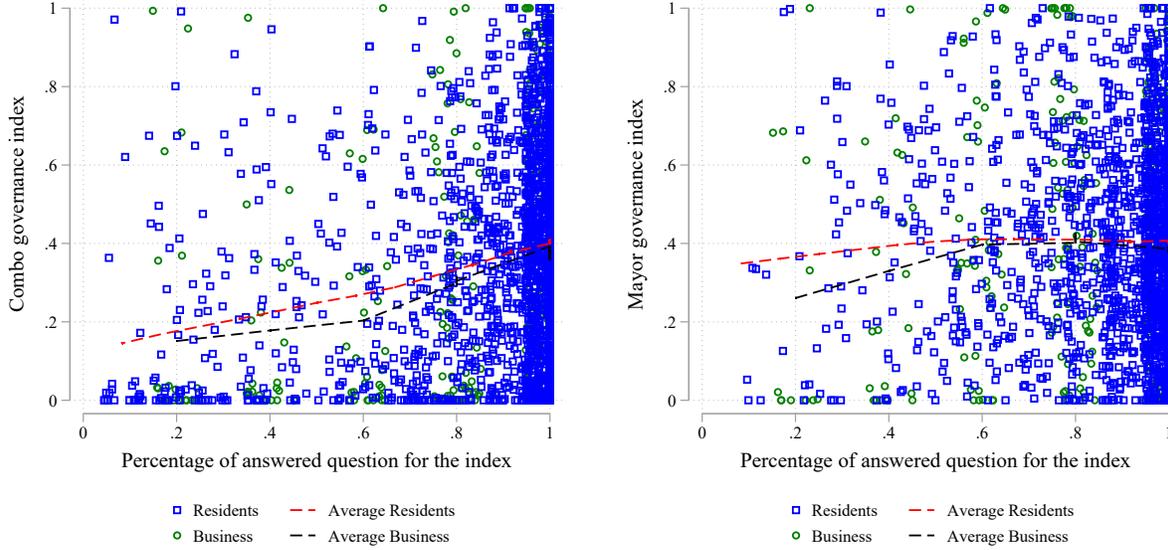
Figure C.1: Difference between randomized response (RR) and direct response (DR) to survey questions on combo “security fee” payment



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.

group, though neither coefficient is statistically significant (see Table C.1).

Figure C.2: Correlation of respondent’s answer percentage and governance levels



Notes: The survey asked about 17 forms of combo and state governance, and the figure plots the percentage of questions each survey respondent answered of these questions by the barrio-level average of combo and state governance.

Table C.1: Treatment-control differences in potential indicators of measurement error

	Control Mean (1)	ATE (2)	SE (3)	P-value (4)	N (5)
Extortion rates					
Sector vacuna payment rate difference (RR-DR)	0.041	-0.050	0.063	0.430	80
Proportion of questions answered					
Proportion of questions answered for relative state governance index	0.033	-0.007	0.009	0.425	80
Proportion of questions answered for mayor governance index	0.891	0.002	0.011	0.865	80
Proportion of questions answered for combo governance index	0.858	0.009	0.013	0.471	80
Proportion of questions answered for relative state legitimacy index	0.082	-0.007	0.011	0.513	80
Proportion of questions answered for state legitimacy index	0.971	-0.011	0.007	0.123	80
Proportion of questions answered for combo legitimacy index	0.889	-0.003	0.013	0.792	80

Notes: This table takes the proxies for measurement error discussed in Appendix C and calculates the correlation with our randomized treatment on these proxies, using the same estimation for our main treatment effects. The vacuna rate difference computes the difference between randomized response and direct response to the question of whether the household pays vacunas. The other measures capture non-response to sensitive items (the proportion of questions answered). We look at the proportion of questions answered for each index, and whether this is different for the state versus the combo. More questions answered for the state could indicate a reluctance to talk about or disclose combo activities.