Mano Dura: An Experimental Evaluation of Military Policing in Cali, Colombia[*]

Robert A. Blair[†]
Michael Weintraub[‡]

Abstract

Governments across the Global South rely on their militaries for domestic policing operations. Advocates argue these policies are necessary to control crime; detractors claim they undermine human rights. We experimentally evaluate a military policing intervention in Cali, Colombia, one of the world's most violent cities. The intervention involved recurring, intensive military patrols randomized at the city block level. Using administrative crime and human rights data, surveys of more than 10,000 residents, a conjoint survey experiment, a costly behavioral measure, qualitative interviews, and firsthand observations from civilian monitors, we find that military policing had no effect on crime during the intervention and adverse effects after it was complete. We also find some suggestive evidence of increased human rights abuses committed by police officers in particular. Despite these null or adverse effects, we find that military policing increased demand for more aggressive military involvement in law enforcement and other aspects of governance.

[*]*Acknowledgements:* Funding for this project was provided in part by the UK Foreign, Commonwealth, & Development Office, awarded through Innovation for Poverty Action's Peace & Recovery Program . Funding for this project was also provided in part by the Open Society Foundations, awarded through Innovation for Poverty Action's Peace & Recovery Program. We thank staff at IPA Colombia (Kyle Holloway and Sofía Jaramillo) for coordination of the project; the Secretaría de Seguridad de Cali (Mará Alejandra Arboleda, Pablo Gracia, Alberto Sánchez, Juan Diego Tabares, Pablo Uribe, and Andrés Villamizar) for logistical oversight; the Third Battalion of the Armed Forces of Colombia (especially Colonel Omar Arciniegas) for its commitment to the evaluation; and Carlos Bohm, María Fernanda Cortés, Abraham Farfán, Sergio Flórez, Carlos Gúzman, Sebastián Hernández, and Lucá Mendoza Mora for terrific research assistance. For helpful feedback we thank Peter Aronow, Graeme Blair, Chris Blattman, Alex Coppock, Erica De Bruin, Leopoldo Fergusson, Don Green, Andrés Moya, Molly Offer-Westort, Matt Ross, Sarah Parkinson, Santiago Tobón, Juan Vargas, Jess Zarkin, and participants at workshops and conferences convened by the American Political Science Association, the American Economic Association, Brown University, Cornell University, Duke University, the Folke Bernadotte Academy, Harvard University's Belfer Center, the Households in Conflict Network, the International Studies Association, Oxford University, Peace Research Institute Oslo, Universidad de los Andes, Universidad del Valle, University of California–Berkeley, University of Pennsylvania, Uppsala University, and the Virtual Crime Economics Online Seminar.
[†]Arkadij Eisler Goldman Sachs Associate Professor, Department of Political Science and Watson Institute for International and Public Affairs, Brown University, robert_blair@brown.edu
[‡]Associate Professor, School of Government, Universidad de los Andes, mlw@uniandes.edu.co

Scholars have long argued that a strict separation between the military and the police is a defining feature of modern democratic regimes (Giddens 1987; Huntington 1957; Kraska 2007). Yet across the Global South, democratically elected governments routinely deploy their militaries for domestic law enforcement. Nowhere is this strategy more common than in Latin America, the world's most violent region (Pion-Berlin and Carreras 2017). In Bolivia, Brazil, Colombia, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Paraguay, and Peru, the armed forces now play a prominent role in domestic policing operations (Flores-Macías and Zarkin 2021*b*). Nor is this strategy unique to Latin America: South Africa, for example, has relied on soldiers to reinforce the police for decades. Military policing is similarly common in Indonesia and the Philippines, among other countries.

Support for these "*mano dura*" (iron fist) policies is widespread, especially in Latin America. Advocates believe military policing is necessary to reduce crime and improve citizens' perceptions of safety (Pion-Berlin and Carreras 2017). Opponents counter that it undermines human rights, delegitimizes the police, and inculcates illiberal and anti-democratic attitudes (Muggah, Garzón and Suárez 2018). Yet despite the increasing prevalence of military policing across the Global South, empirical evidence of its efficacy is surprisingly scarce. Because soldiers are typically only deployed when crime is rampant and police forces are overwhelmed, observational studies of military policing confront severe inferential obstacles, and much of the debate remains anecdotal and impressionistic.[1]

We report results from an experimental evaluation of the *Plan Fortaleza* military policing program in Cali, Colombia, the country's third largest city and one of its most violent. In 2018, Cali reported a homicide rate of 46.7 per 100,000 residents, nearly double that of Colombia's second largest city (Medellín) and more than triple the capital's (Bogotá's). In response, the government of Cali deployed recurring, intensive military patrols to two *comunas* (communes) with some of the highest homicide rates in the city. Military policing programs like *Plan Fortaleza* are a defining characteristic of *mano dura* policies in Latin America and beyond (Holland 2013,

[1]For recent exceptions, see Espinosa and Rubin (2015); Flores-Macías (2018); Flores-Macías and Zarkin (2022); Flores-Macías and Zarkin (2021*b*).

46). We partnered with the Mayor's Office, the Third Brigade of the Colombian Armed Forces, and Innovations for Poverty Action (IPA) Colombia to experimentally evaluate the impact of these operations.

We develop a theoretical framework to explain the effects of military policing on the behavior of four sets of actors: criminals, civilians, police officers, and soldiers. Following our pre-analysis plan (PAP),[2] we then test the observable implications of our theory by evaluating the effects of *Plan Fortaleza* on three sets of outcomes: security, human rights abuses, and political attitudes and behaviors. We combine administrative data on crime and human rights abuses with two waves of surveys reaching over 10,000 respondents in total; a conjoint survey experiment; in-depth, semi-structured qualitative interviews with 49 civil society leaders; and detailed firsthand observations from civilian monitors hired to accompany the soldiers while on patrol. We also created a hotline that residents could text or call to petition for additional military presence in their neighborhoods—a "costly" behavioral measure of demand for military policing to help address potential social desirability concerns.

We find little to no evidence that *Plan Fortaleza* reduced the objective prevalence of crime while the intervention was ongoing, and if anything it appears to have *exacerbated* crime after the intervention was complete. We similarly find little to no evidence that the program improved citizens' subjective perceptions of safety. We find some evidence of increased human rights abuses, albeit only in the survey data. In almost all cases these abuses appear to have been perpetrated by police officers rather than soldiers, likely in the course of making arrests. We find no evidence of increased abuses by either police officers or soldiers in data from the Attorney-General's Office or in the firsthand accounts of civilian monitors.

Yet despite these null or adverse effects on security and human rights, we find that *Plan Fortaleza* increased demand for more aggressive military involvement in law enforcement and other aspects of governance. Residents who were exposed to the program expressed more favorable perceptions of the military, and were more likely to call or text the hotline to request more mil-

[2] Our PAP was preregistered with the Evidence in Governance and Politics (EGAP) network prior to endline survey data collection. We include an anonymized version of the PAP as an appendix.

itary patrols. While surprising, these results are consistent with previous studies (Flores-Macías and Zarkin 2022), and suggest that civilians' support for military policing may be inelastic to its efficacy. We return to this result in the discussion. More alarmingly, we find that residents who were exposed to military policing were also more likely to express support for vigilantism and for military coups in response to rising crime. These results suggest the possibility of a weakening commitment to democracy and the rule of law among the program's intended beneficiaries.

Finally, we exploit municipal elections held in the middle of the intervention to test the impact of military policing on residents' political preferences. Scholars of "penal populism" argue that right-wing politicians use dubious *mano dura* policies to win votes (Holland 2013; Pratt 2007). Contrary to these accounts, we find no evidence that *Plan Fortaleza* increased turnout for right-wing candidates. Through a desk review of the candidates' policy platforms and qualitative interviews with civil society leaders, we show that right- and left-wing politicians in Cali are generally perceived to be equally "tough on crime," and equally supportive of military policing. This helps explain the null effect on vote choice. We also exploit massive nationwide protests that occurred immediately after the intervention to test the program's impact on support for civil disobedience associated with left-wing unions, indigenous groups, and student organizations. We find that residents who were exposed to military policing were weakly less supportive of the protests, though no more or less likely to participate in them.

Our study contributes to research on hot spots (or "place-based") policing (Blattman et al. 2021; Collazos et al. 2019; Braga and Bond 2008; Weisburd and Telep 2014), penal populism (Holland 2013; Pratt 2007), and the militarization of law enforcement in Latin America and beyond (Flores-Macías and Zarkin 2021*a*). Most studies of militarization focus on the impact of SWAT team deployments or military hardware transfers to police departments in the US, with mixed results (Bove and Gavrilova 2017; Delehanty et al. 2017; Harris et al. 2017; Gunderson et al. 2021; Lowande 2021; Mummolo 2018). Other studies address militarization of the police globally (De Bruin 2022) or in Latin America specifically (González 2020; Magaloni and Rodriguez 2020; Magaloni, Franco-Vivanco and Melo 2020).

In contrast, relatively few scholars have tested the effects of military policing itself—i.e. deployments of soldiers to conduct domestic policing operations. This gap is important not just because military policing is extremely popular (Pion-Berlin and Carreras 2017) and increasingly widespread (Flores-Macías and Zarkin 2021*a*), but also because militaries differ from even the most militarized police forces in the type and intensity of training they receive; in their more inflexible vertical command structures; and in the expectation that they will use physical and often lethal force to eliminate enemies on the battlefield. Only a handful of studies have explored the implications of military policing for the rights and safety of civilians, all using observational data (Espinosa and Rubin 2015; Flores-Macías 2018; Flores-Macías and Zarkin 2022). To our knowledge, ours is the first experimental evaluation of this increasingly common but still poorly understood approach to law enforcement.

1 Theoretical framework

Our theoretical framework consists of four sets of actors: criminals, civilians, police officers, and soldiers. We seek to understand the implications of military policing for each. In particular, we explore whether and how military policing affects (1) the probability that criminals will commit crimes; (2) the probability that police officers and soldiers will perpetrate human rights abuses; and (3) civilians' perceptions of safety and their attitudes and behaviors towards police officers, soldiers, and the criminal justice system more generally. Because the intervention we study was randomized at a very local level, we assume the context in which these actors operate is identical in expectation *except* for the presence of the military, allowing us to abstract away from the various other structural and environmental factors (e.g. poverty) that may affect our outcomes of interest.

1.1 Security

Classic deterrence theories hold that criminals are rational actors whose decision to commit crime depends on the certainty, severity, and swiftness of punishment (see Braga et al. 2019 for a review).

Advocates of military policing draw on these theories to argue that soldiers are more effective than police officers at ensuring certain, severe, and swift punishment for criminals. Militaries typically have better communication, transportation, and logistical capacity than police forces, and soldiers typically receive more (and more rigorous) training than police officers, especially in Latin America (Flores-Macías 2018; Kraska 2007). According to proponents, these distinctions imply that military policing should increase the perceived certainty and swiftness of punishment. (We consider reasons to be skeptical of this claim below.)

Equally important, advocates argue that soldiers are less prone to corruption than police officers. Even in countries where soldiers play an active role in domestic law enforcement, reported rates of corruption are higher for the police than the military (Pion-Berlin and Carreras 2017). Police officers enjoy discretion when deciding whether and how to enforce particular laws, which can exacerbate corruption; soldiers, in contrast, are expected to follow orders with less autonomy and under stricter hierarchies of command and control, which should, in principle, mitigate corruption (Campbell and Campbell 2010, 339). (Again, we consider reasons to be skeptical of this proposition below.) If criminals believe soldiers will not exchange bribes for leniency, then military policing should increase the perceived certainty of punishment.

Perhaps most obviously, soldiers also have more sophisticated weapons, munitions, and other forms of coercive capacity than police officers, and are trained to capture or eliminate enemy combatants on the battlefield. If criminals believe soldiers will kill or injure them in a confrontation, then military policing should increase the perceived severity of punishment, regardless of whether an arrest is made. The military's coercive capacity may threaten human rights—a possibility we discuss in further detail below. But it may also improve deterrence. Again, these distinctions are likely to be especially relevant in Latin America, where many police forces are overwhelmed by "daunting" organized criminal syndicates (Zechmeister 2014, 104), and where some have been shown to be complicit with organized crime (Flores-Macías and Zarkin 2021*a*, 5).

Finally, if civilians share the expectation that soldiers will be more effective than police officers at deterring criminals, then military policing should not only reduce the objective prevalence of crime, but should also improve subjective perceptions of safety. In many countries in the Global South, and in Latin America in particular, citizens tend to perceive the military as more effective and better trained than the police (Pion-Berlin and Carreras 2017). They also tend to express greater trust in the military; indeed, other than the Catholic Church, no other institution is as widely trusted as the military in the region (Zechmeister 2014, 114).

Following the discussion above, in our PAP we hypothesized that:

H1: Military policing will reduce the objective prevalence of crime.

H2: Military policing will improve civilians' subjective perceptions of safety.

1.2 Human rights

Critics of military policing argue that it induces soldiers (and possibly police officers as well) to violate the human rights of criminals and civilians. Reliance on the military for domestic policing operations replaces the "traditional police orientation of 'protect and serve"' with a "military orientation of 'overwhelm and defeat"' (Campbell and Campbell 2010, 329-30). Soldiers may be "hard-wired" to use force in ways that police officers are not (Pion-Berlin and Carreras 2017, 9). Worse, police officers who serve alongside soldiers may imitate their abusive behavior—a diffusion of social norms born from police officers' tendency to perceive militarized policing strategies as more "elite" (Kraska and Kappeler 1997). In this way, military policing may inadvertently disseminate policing practices that are "intrinsically incompatible" with protecting and serving civilians (Greener-Barcham 2007, 93).

Proponents of military policing counter that soldiers may be *less* rather than more prone to human rights abuses than police officers—a belief shared by a majority of Latin American citizens (Pion-Berlin and Carreras 2017). The standards governing police use of force are weaker than those governing the military in many countries, while the consequences for transgressing use of force policies are less severe for police officers than for soldiers (Tecott and Plana 2016; Wood 2015). While impunity is common in Latin America, and while many of the region's militaries have reputations for abuse, the same or worse is often true of Latin American police forces. As Pion-Berlin and Carreras (2017, 7) explain, assassinations, kidnappings, and torture have become "familiar police practices" throughout the region. A lack of training and resources has only exacerbated these problems (Brinks 2007).

In our PAP we therefore hypothesized that:

Hypothesis 3: Military policing will have at worst no effect on human rights abuses.

1.3 Political attitudes and behaviors

Military policing may also have profound and lasting effects on civilians' political attitudes and behaviors. Most directly, if military policing reduces the objective prevalence of crime and improves subjective perceptions of safety without increasing human rights abuses, it may legitimize the military at the expense of the police. When citizens express support for military involvement in domestic law enforcement, they make a "comparative judgment call about the relative efficacy of military versus police conduct in domestic security roles" (Pion-Berlin and Carreras 2017, 5). If soldiers are more effective and less abusive than police officers, then military policing may bolster already favorable attitudes towards the military while further eroding already unfavorable attitudes towards the police. This may also create demand for additional military policing in the future.

In our PAP we hypothesized that:

Hypothesis 4: Military patrols will increase the perceived legitimacy of the military in the eyes of civilians.

Hypothesis 5: Military patrols will decrease the perceived legitimacy of the police in the eyes of civilians.

Military policing may also generate support for military intervention in other areas of governance, potentially eroding citizens' commitment to democracy and the rule of law. Political parties born of military dictatorships often promote military policing and other *mano dura* policies to invoke memories of a time when crime was less pervasive, using the "language, figures, and founding myths" of an authoritarian past to demonstrate their commitment to security "at all costs" (Holland 2013, 52). Citizens exposed to military policing may become more receptive to these myths, concluding that security justifies violations of due process protections (e.g. the use of vigilantism to punish offenders) and broader liberal democratic values (e.g. the establishment of military governments to control crime).

We therefore hypothesize that:

> **Hypothesis 6:** Military policing will weaken civilians' commitment to democracy and the rule of law.[3]

Finally, the effects of military policing on civilians' political attitudes and behaviors may spill over onto the parties and politicians with whom it is typically associated. Military policing is a form of penal populism, in which elected officials adopt *mano dura* policies to satisfy voters' demand for law and order (Pratt 2007). In Latin America, penal populism is generally associated with the right wing (Holland 2013; Visconti 2020)—though, as we discuss below, politicians from across the political spectrum have supported military policing in Colombia, our study's setting. The logic of penal populism suggests that military policing may increase citizens' support for right-wing policies, parties, and politicians, and reduce their support for left-wing alternatives.

We hypothesize that:

> **Hypothesis 7:** Military patrols will increase support for right-wing causes and candidates and decrease support for left-wing alternatives.[4]

[3] This hypothesis deviates somewhat from our PAP, where we hypothesized that military policing would increase support for retributive justice. Here we broaden this hypothesis to encompass commitment to democracy and the rule of law more generally, as we think this outcome is of greater political and normative concern. We find no evidence that *Plan Fortaleza* affected support for retributive justice specifically.

[4] Again, this hypothesis deviates somewhat from our PAP. In our PAP we hypothesized that military policing would increase turnout for right-wing candidates. Here we broaden this hypothesis to encompass support for right-wing causes more generally.

1.4 Reasons for skepticism

As we note in our PAP, none of these hypotheses is uncontroversial. While we hypothesized that military policing would reduce the objective prevalence of crime, soldiers may prove less effective than police officers at deterring criminals. For example, deterrence may require the ability to gather evidence and investigate suspects (Bayley 1998)—something soldiers are not trained to do. Organized criminal groups may also respond strategically to the threat of military force, generating more rather than less crime (Lessing 2015). Even in the best cases, deterrence may depend on soldiers' physical presence on the streets, such that crime falls temporarily at the times and places that soldiers are active, only to rise again when and where they are not (e.g. Blattman et al. 2021).

We hypothesized that military policing would improve civilians' subjective perceptions of safety. But citizens may interpret increased military presence as a signal that their neighborhoods are unsafe, exacerbating perceptions of insecurity (Rosenbaum 2006). They may also fear being stigmatized by soldiers, especially if they belong to historically marginalized groups (Dammert and Bailey 2007). Finally, we hypothesized that military policing would have at worst no effect on human rights abuses. But the military's focus on "overwhelming and destroying an adversary" may create tradeoffs between deterrence and human rights (Campbell and Campbell 2010, 329-30). As we discuss below, we developed detailed protocols to monitor and respond to any adverse effects of the *Plan Fortaleza* program while it was in the field. But there is very little empirical evidence to adjudicate between the competing predictions above, and our hypotheses are not foregone conclusions. They are empirical propositions in need of rigorous testing.

2 Setting and intervention

Colombia has one of the highest crime rates in Latin America. We focus on the city of Cali, perhaps best known as the home of the infamous Cali Cartel, which controlled much of the trafficking and transshipment of cocaine from Colombia in the 1980s and 90s (Vanegas Muñoz 2021). Organized crime is weaker and more fragmented today, more focused on "micro-territories," and more dedicated to money laundering, extortion, and micro-trafficking of drugs for domestic consumption. Violence is driven primarily by disputes between gangs. Despite improvements in recent decades, crime remains pervasive, and Cali ranks among the most violent cities in the world.[5]

To address these problems, Cali's Security and Justice Secretariat deployed the military to patrol crime hot spots as part of an initiative known as *Plan Fortaleza*. While this initiative is specific to Cali, it is part of a trend towards increasing military involvement in domestic law enforcement throughout Colombia. In 2016, the government signed a peace agreement with the *Fuerzas Armadas Revolucionarias de Colombia* (FARC), the country's largest rebel group. With the FARC's demobilization, the military began to shift from counterinsurgency in the countryside to domestic policing operations in cities, including Bogotá, Medellín, Barranquilla, and, most prominently, Cali. Cali is thus an important test case for military policing in Colombia and in Latin America more generally, where military policing has been widely adopted but seldom evaluated. We discuss generalizability in further detail in the conclusion.

Plan Fortaleza consisted of recurring, intensive vehicular and foot patrols by heavily armed soldiers from two units of the Armed Forces of Colombia: the Military Police and the Special Forces. The Special Forces tend to be older, more experienced, and better equipped than the Military Police. Cali has 22 *comunas* in total; *Plan Fortaleza* focused on *comunas* 18 and 20, hot spots for crime with a combined population of approximately 215,000 residents. To minimize logistical problems, the Military Police and Special Forces never patrolled the same *comuna* on the same day, instead alternating on a 12-day rotation. Patrols were conducted by teams of six to eight soldiers, with seven to eight teams patrolling more or less simultaneously between the hours of 5:00pm and midnight every weekday night. While on patrol, soldiers searched residents for possession of drugs and weapons ("*requisas*"), checked IDs and business licenses, erected road blocks, detained suspected criminals, and conversed with residents.

We randomly assigned patrols at the *manzana* (city block) level. Each treatment block was assigned to receive approximately 30 minutes of military patrols once every six days. In reality,

[5] According to data from the NGO *Seguridad, Justicia, y Paz*, Cali had the 21st highest homicide rate in the world at the time of our study.

the average time spent patrolling was around 11 minutes per block per day, due to the small size of most blocks[6] and the large number of soldiers on patrol. All blocks also had some police presence, albeit limited. Importantly, as we show in Table A.11 in the Appendix, we observe no difference in police presence between treatment and control blocks, suggesting the police did not displace from the treatment group to the control group (or vice versa) as a result of the intervention. Our evaluation began on September 30, 2019 and concluded on November 18, 2019, when massive nationwide protests required a redeployment of the military to other sites.

While the intervention was relatively short, it was also arguably quite strong, with large teams of soldiers patrolling small city blocks carrying assault rifles and wearing fatigues, helmets, and flak jackets. Patrols were scheduled for times when civilians tend to be at home or in the streets, making them more likely to observe the soldiers' presence. While Colombian cities have pursued increasingly militarized approaches to policing in recent years, the sight of soldiers patrolling residential neighborhoods remains a rarity. Moreover, the length of the program was not atypical of military policing interventions in Latin America, where soldiers usually participate in temporally and geographically delimited operations targeting hot spots for crime and/or drug trafficking (Flores-Macías and Zarkin 2021*a*, 526). Longer military occupations are less common, though they do sometimes occur (Flores-Macías and Zarkin 2021*b*).

Comunas 18 and 20 are densely populated and difficult to navigate. Civilian monitors accompanied the soldiers, using GPS devices and smartphones equipped with a customized Google Maps interface to help them locate their assigned treatment blocks and to monitor their activities while on patrol. For safety reasons, the monitors were provided bulletproof vests and were required them to stay in their vehicle at all times. At no point during the evaluation did we receive any reports of attempted or actual violence against monitors. We provide descriptive statistics on the patrols in Table A.3 in the Appendix, and discuss our safety and ethical protocols, and the possibility of Hawthorne effects induced by the monitors' presence, in further detail below.

[6]The perimeter of the average block in our sample is 283 meters, with a standard deviation of 248 meters.

3 Research Design

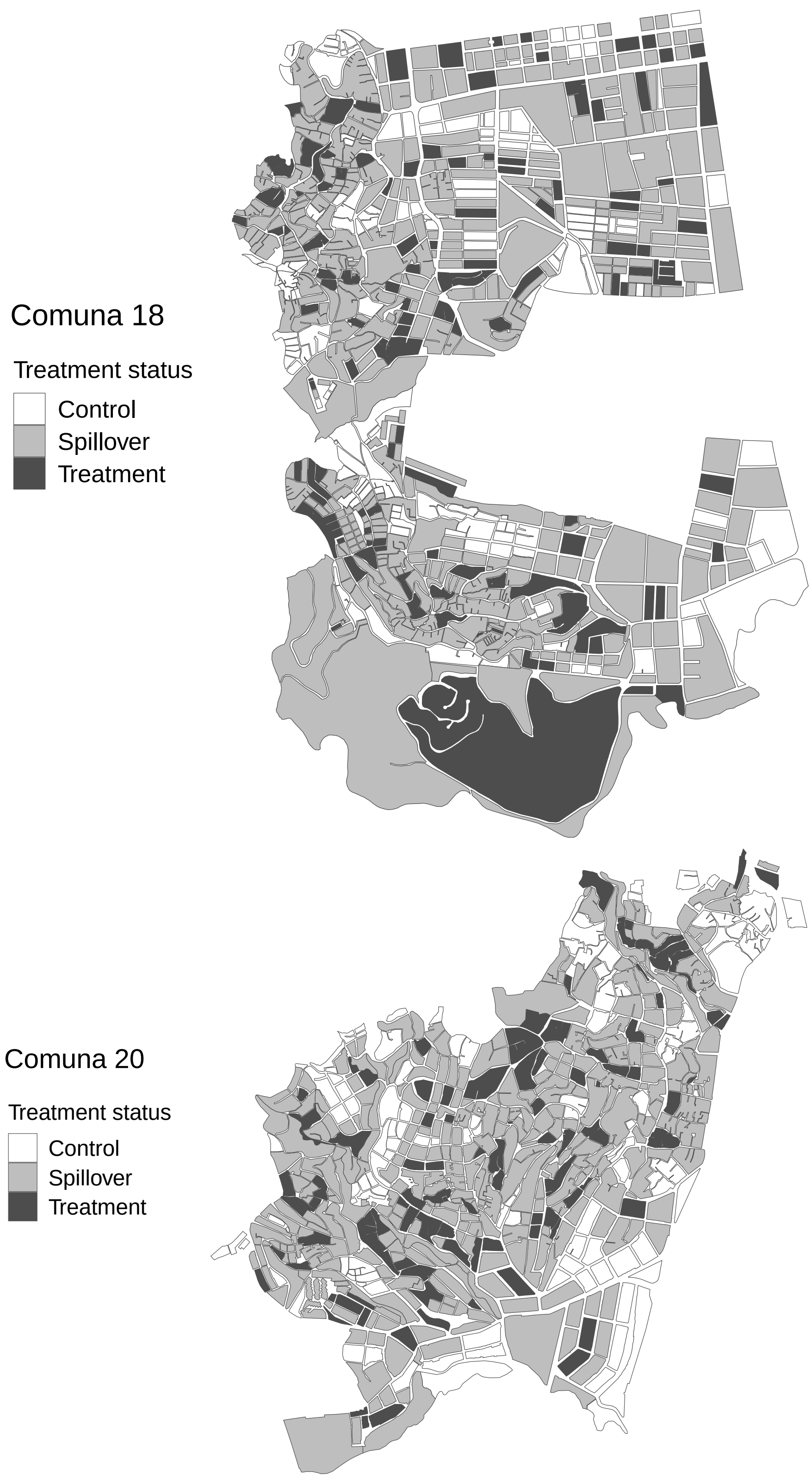
3.1 Randomization

The two *comunas* in our sample comprise 1,255 blocks nested within 30 *barrios* (neighborhoods). We stratified by neighborhood, then randomly assigned 1/6 of all blocks in each neighborhood to treatment. (We gain no additional statistical power for estimating treatment effects, and only marginal statistical power for estimating spillover effects, when increasing the proportion of treated blocks beyond 1/6. We provide power calculations in Appendix A.1.) We recognized at the outset that the probability of spillover would be high in these densely populated neighborhoods, and so assigned to the spillover group any block that (1) was adjacent to at least one treatment block but (2) was not itself assigned to treatment, following the procedure described in Appendix A.2. We assigned all remaining blocks to the control group. We thus make the simplifying assumption that the effects of the program spillover from treatment to adjacent blocks, but no further. We test for more complex spillover dynamics, including saturation and linear and exponential decay, in Appendix A.9.1. Figure 1 maps the blocks in our sample by treatment assignment.

3.2 Data

We combine data from five sources. First, we collected administrative data on crime, timestamped and geocoded to the block level. These data begin nine months before the intervention (January 1, 2019) and end six weeks after (December 31, 2019). The quality of administrative crime data is unusually high in Cali, where representatives from the Mayor's Office meet regularly with the Colombian National Police, the Attorney-General's Office, and the Coroner's Office to approximate the true prevalence of homicides and other violent crimes. We also collected administrative data on human rights abuses over the same period from the Office of the Attorney-General, which is responsible for investigating and prosecuting most cases of misconduct by the police and military. These data include alleged abuses reported by victims and witnesses, timestamped and geocoded

Figure 1: **Treatment assignments**



to the block level.

Second, we conducted a survey of 2,096 randomly selected residents and business owners between October 17 and December 19, 2019, beginning while the intervention was ongoing and continuing for roughly a month after it concluded. We surveyed three residents and two business owners on each of 416 randomly selected blocks: 202 from the treatment group, 109 from the spillover group, and 105 from the control group. We over-sampled treatment blocks in order to monitor treatment compliance and document abuses while the soldiers were on patrol. We refer to this as the monitoring survey.

Third, we conducted another survey of 7,921 randomly selected residents and business owners on all blocks between January 17 and February 25, 2020, between two and three months after the end of the intervention. We surveyed six residents and two business owners per block, following the sampling frame described in Appendix A.5. We refer to this as our endline survey. The survey also included a forced-choice, image-based conjoint experiment inspired by Flores-Macías and Zarkin (2022) and a costly behavioral measure involving a hotline that respondents could text or call to request additional military patrols in their neighborhood, described in further detail below. There were five blocks (0.3% of the sample) where it was impossible to administer our endline survey due to safety concerns, and another three where it was impossible to implement our costly behavioral measure. We drop these blocks from our analysis.

Fourth, we collected GPS data and detailed firsthand observations from the civilian monitors hired to accompany the soldiers while on patrol. We use these data to measure patrol duration, the number of soldiers on each patrol, and the soldiers' activities, including any acts of verbal or physical abuse. (As we discuss below, the monitors recorded no incidents of physical abuse and only one minor incident of verbal abuse over the course of the evaluation.) Finally, we conducted semi-structured, in-depth qualitative interviews with 49 civil society leaders from *comunas* 18 and 20, selected using snowball sampling. We use these interviews to better understand the program's effects on political beliefs and behaviors, as discussed below.

3.3 ESTIMATION

Most of our outcomes are measured at the individual level using survey data; outcomes based on administrative or behavioral data are measured at the block level instead. We estimate the intention-to-treat (ITT) effect of the program using a weighted least squares (WLS) regression where observations are weighted by the inverse probability of assignment to their realized treatment status. Because the probability of assignment to the spillover and control groups depends on proximity to the nearest treatment block, these weights cannot be computed analytically. Instead, we bootstrap our randomization procedure and estimate the probability that each block is assigned to each treatment condition across 1,500 replications.

When testing the effects of the program on individual-level outcomes, we fit a WLS regression given by

$$y_{ijk} = \theta t_{jk} + \lambda s_{jk} + \beta \mathbf{X_{jk}} + \delta \mathbf{Z_{ijk}} + \alpha_k + \epsilon_{ijk} \tag{1}$$

where y_{ijk} denotes the outcome for respondent i on block j in neighborhood k; t_{jk} denotes assignment to treatment; s_{jk} denotes assignment to spillover; $\mathbf{X_{jk}}$ denotes block-level covariates;[7] $\mathbf{Z_{ijk}}$ denotes individual-level covariates;[8] α_k denotes neighborhood fixed effects; and ϵ_{ijk} is an individual-level error term, clustered by block. When testing the ITT on block-level outcomes, we omit the individual-level covariates and do not cluster our standard errors.

Because of the way blocks are distributed across neighborhoods, some (though very few) have 0 probability of assignment to the spillover or control group.[9] We exclude blocks with 0 probability of assignment to control when estimating the ITT, following Aronow and Samii (2017). When estimating spillover effects, we exclude blocks with 0 probability of assignment to control as well as blocks with 0 probability of assignment to spillover. We discuss deviations from our

[7] We control for area of the block, distance to the nearest police station and military battalion, and distance to the nearest public transportation hub based on administrative data. We also control for the average age, average years of education, and percentage of men on each block, aggregating our individual-level survey data up to the block level. We include a lagged dependent variable when testing the effect of the intervention on crime in the administrative data.

[8] We control for age, gender, years living on the block, and, when available, years of education. (We measured years of education in the endline but not the monitoring survey.)

[9] Of the 1,255 blocks in our sample, one had 0 probability of assignment to spillover and another three had 0 probability of assignment to control.

PAP in Appendix A.7, and report results with multiple comparisons corrections in Appendix A.10.

3.4 Spillover and "fuzzy clustering"

Spillover effects raise a related concern about what Blattman et al. (2021) call "fuzzy clustering." The problem arises because the probability of spillover in our sample is a function of geographical proximity, creating dependencies that span administrative units (e.g. neighborhoods). While this problem is endemic to studies of place-based policing—and, indeed, to almost all studies in which geographical spillover is at least plausible—methodologists disagree about how and under what conditions to adjust for it (Abadie et al. 2023; Aronow and Samii 2017). Following Blattman et al. (2021), in Appendix A.10 we report randomization inference (RI) p-values for all ITT estimates in the paper. RI tests the sharp null hypothesis of no treatment effect for any unit. While RI is controversial (Gelman 2003; Keele 2015), it provides a useful complement to more conventional test statistics under arbitrarily complex forms of spatial interference.

3.5 Compliance

To track treatment compliance, we established geo-fences of 25 meters around each treatment block and calculated the amount of time that each patrol spent within its assigned geo-fence using GPS data collected by the civilian monitors. On any given night, soldiers correctly patrolled between 85% and 100% of their assigned treatment blocks. Further corroborating the GPS data, in Table A.11 we show that, as expected, residents of treatment and spillover blocks were much more likely to report military presence than residents of control blocks, and that the magnitude of the effect is larger on treatment blocks than on spillover blocks.

3.6 Hawthorne effects

The presence of civilian monitors may have caused the soldiers to modify their behavior, inducing something akin to a Hawthorne effect. To mitigate this risk, monitors were instructed to remain in their vehicle while on patrol, and were trained to be discreet and unobtrusive when observing soldiers' actions. To facilitate discretion and standardize data collection, monitors used a smartphone app developed for this project to record their observations. Because monitors accompanied the patrols every weekday night for nearly two months, the soldiers likely grew inured to their presence (Blair, Karim and Morse 2019). Moreover, due to the density of these neighborhoods and the relative novelty of military patrols, it is likely that residents would have watched the soldiers' actions—and that the soldiers would have known they were being watched—even without the monitors. Nonetheless, our results should be interpreted with this caveat in mind.

3.7 Ethics

Given the increasing prevalence of military policing in Colombia and across the Global South; the lack of evidence on its efficacy; and the arguments of advocates (including both residents and elected officials in Cali and elsewhere) that military policing is necessary to combat crime, we believed a rigorous impact evaluation was needed to inform both theory and policymaking. Given the troubled history of the Colombian police, there were reasons to believe military policing might deter criminals and reduce human rights abuses. Moreover, the *Plan Fortaleza* program predated our study, and would have continued with or without our evaluation of it. Municipal authorities and military officials had already selected the *comunas* and neighborhoods in which the intervention would occur. We randomized only the blocks that soldiers would and would not patrol.

Nonetheless, both the program and our evaluation of it posed risks, which we sought to anticipate and mitigate. We summarize our precautions here, and discuss them in more detail in Appendix A.6. First, there was a risk that military patrols would expose residents to human rights abuses. We sought to minimize this risk by using the monitors' observations to document and report any abuses to the Security and Justice Secretariat, which is responsible for overseeing the military's operations in the city. The monitors recorded only one minor incident of verbal abuse involving a soldier throughout the duration of the study. We further corroborated the monitors' reports with data from our monitoring survey and the Attorney-General's Office.

Second, there was a risk that military patrols would subject residents to violence by shifting the equilibrium distribution of gang presence and activity. We determined that this risk was minimal. Our conversations with the military and the Security and Justice Secretariat strongly suggested that such an equilibrium did not exist at the time of our study, given the city's highly fragmented landscape of organized crime. Prior to our study, the military (non-randomly) varied its patrol routes from day to day to prevent criminals from adapting; it continued this practice during the evaluation, for the same reason. This should have further reduced the risk of a change in the equilibrium distribution of gang presence and activity.

Third, there was a risk that respondents would face reprisals for participating in our surveys, or that enumerators would face reprisals for administering them. To minimize this risk, enumerators received specialized training and followed strict security protocols, including a requirement that they complete data collection by noon each day. All surveys were conducted in private, and respondents were repeatedly informed that their participation was voluntary and anonymous; that the survey could be halted at any time; and that they could skip any question they did not want to answer. Both before and during data collection, we consulted local IPA staff, field supervisors, and civil society representatives to diagnose whether particular blocks posed especially acute security concerns, and we adjusted our data collection protocols accordingly. There were no reports of threats or violence against respondents or enumerators at any time during the evaluation.

Fourth, there was a risk that criminals would recognize and threaten the civilian monitors. To minimize this risk, we recruited monitors who did not live in the *comunas* in our sample, reducing the probability that they would be identified. Monitors also maintained a direct line of communication with the military and the Security and Justice Secretariat, which they could use to seek help if necessary. As discussed above, monitors were instructed to remain in their vehicles at all times, and were provided with bulletproof vests and armbands clearly identifying them as civilians to minimize the risk that they would be mistaken for soldiers and attacked. There were no reports of threats or violence against monitors at any time during the evaluation.

4 Results

4.1 Security

Null or adverse effects on objective prevalence of crime and subjective perceptions of safety

In Table 1 we report the ITT of the *Plan Fortaleza* program on crime and perceptions of safety. The dependent variables in columns 1 and 2 are additive indices of murders, armed robberies, thefts, illegal drug sales, and illegal possession of firearms based on administrative data; we distinguish crimes committed during the intervention (column 1) from those committed between the end of the intervention and the end of the year (column 2).[10] In columns 3 and 4, the dependent variable is a standardized additive index of victimization by vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, domestic violence, and extortion based on the endline survey. Respondents were asked whether and when they or someone in their household had been victimized by any of these crimes in the past six months; we again distinguish crimes committed during the intervention from those committed after.

To construct the dependent variable in column 5, respondents were asked how frequently they had witnessed each of 15 crimes in the past month on a four-point Likert scale: the 10 listed above, as well as prostitution, illegal drug sales, illegal drug consumption, public alcohol consumption, and illegal possession of firearms. We aggregate answers to these questions into a standardized additive index. Because respondents were asked about crimes witnessed in the past month, and because the endline survey was conducted in January and February of 2020, we assume that any crimes witnessed likely occurred after the intervention ended in November 2019.

Contrary to hypothesis 1, we find no evidence that *Plan Fortaleza* reduced the objective prevalence of crime during the intervention (columns 1 and 3), and if anything we find that it

[10] In our PAP we proposed weighting crimes in the administrative data by the average prison sentence under Colombian law. Based on feedback from criminologists, we determined that this approach is atypical and yields results that are difficult to interpret. We report results using unweighted indices instead.

Table 1: **Treatment effects on crime, crime victimization, crime witnessing, and perceptions of safety**

	Admin data		**Survey data**			
	Crime incidence		**Crime victimization**		**Crime witnessing**	**Safety perceptions**
	During intervention	**After intervention**	**During intervention**	**After intervention**	**After intervention**	**After intervention**
Treatment	0.003	0.110**	0.006	-0.007	0.153***	-0.050
	(0.036)	(0.050)	(0.042)	(0.047)	(0.052)	(0.047)
Spillover	-0.038	0.083*	0.026	0.013	0.186***	-0.066
	(0.030)	(0.044)	(0.030)	(0.038)	(0.043)	(0.040)
Individual controls	✗	✗	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓	✓
Observations	1167	1167	7845	7845	7837	7707
R^2	0.33	0.48	0.03	0.03	0.12	0.09
Control mean	0.160	0.160	-0.021	-0.016	-0.119	0.077

Notes: ITT on additive indices of crime during (column 1) and after (column 2) the intervention based on administrative data, and standardized additive indices of crime victimization during (column 3) and after (column 4) the intervention, crime witnessing after the intervention (column 5), and perceptions of safety after the intervention (column 6) in the endline survey data. All specifications include neighborhood fixed effects and block-level controls. Columns 1 and 2 include a lagged dependent variables; columns 3-6 include individual-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are in parentheses, and are clustered by block in columns 3-6. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

exacerbated crime after the program was complete (columns 2 and 5). Using the administrative data, we observe substantively large increases of 69% more crimes on treatment blocks and 52% on spillover blocks relative to the control group mean after the intervention (0.160 crimes per block). We also observe a 0.153 standard deviation increase in respondents' reports of witnessing crimes on treatment blocks and a 0.186 standard deviation increase on spillover blocks. As we show in more exploratory analyses in Appendices A.11.2 and A.11.3, these increases in crime are accompanied by increases in crime reporting, police presence, and arrests on both treatment and spillover blocks. While we do not find a corresponding increase in endline survey respondents' reports of crime victimization after the intervention (column 4), this may be because randomly selected respondents were more likely to have witnessed crimes than to be victims of them.

To construct the dependent variable in column 6 of Table 1, respondents were asked how safe they feel walking on their block during the day and at night on a five-point Likert scale; how safe they feel using a smartphone on the street; and how worried they are about becoming victims of a violent or non-violent crime in the next two weeks. Residents were also asked about any precautions they had taken for fear of crime in the past month, including avoiding public transportation, staying home at night, or prohibiting children from playing in the street or attending school. We aggregate answers to these questions into a standardized additive index. Contrary to hypothesis 2, we find no evidence that *Plan Fortaleza* improved civilians' subjective perceptions of safety, except possibly among business owners.[11] Taken together, these results suggest that *Plan Fortaleza* had at best no beneficial effect and at worst adverse effects on the outcomes most often touted by proponents of military policing.

[11] In exploratory analyses, we find that the intervention improved subjective perceptions of safety among business owners specifically. But these effects are unique to business owners, and we are careful not to over-interpret them.

Table 2: **Treatment effects on human rights abuses**

	Monitoring data		Survey data	
	Police abuse	**Military abuse**	**Police abuse**	**Military abuse**
Treatment	0.037***	0.010*	0.011	-0.001
	(0.011)	(0.006)	(0.013)	(0.004)
Spillover	0.016	0.001	0.030**	0.002
	(0.011)	(0.003)	(0.012)	(0.003)
Individual-level controls	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓
Observations	1970	1970	7908	7908
R^2	0.05	0.05	0.05	0.01
Control mean	0.015	0.000	0.114	0.012

Notes: ITT on abuses by police (columns 1 and 3) and military (columns 2 and 4) based on monitoring (columns 1 and 2) and survey data (columns 3 and 4). All specifications include neighborhood fixed effects and individual- and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors, clustered by block, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

4.2 Human rights

Weak or adverse effects on human rights abuses, depending on data source

In Table 2 we report the ITT of the program on human rights abuses. To construct the dependent variable in columns 1 and 2, monitoring survey respondents were asked how often they had seen or heard about physical or verbal abuses committed by police officers or soldiers in the past two weeks; for columns 3 and 4, endline survey respondents were asked if they had seen or heard about any abuses in the past month. Verbal abuses included insulting, shouting, cursing, and other forms of verbal assault; physical abuses included pushing, hitting, and other forms of physical assault.

Because the endline survey was conducted in January and February of 2020, we assume that any abuses reported by endline survey respondents occurred after the intervention. While the monitoring survey continued for roughly one month after *Plan Fortaleza* was complete, to allow for telescoping we assume that any abuses reported by monitoring survey respondents may have occurred during the intervention. In Appendix A.11.1 we probe the robustness of our results to different coding rules for distinguishing abuses that occurred during the intervention from those that occurred after. Since most respondents who reported physical abuse also reported verbal abuse, we collapse the two categories into a single indicator.

Contrary to hypothesis 3, we find some suggestive evidence that *Plan Fortaleza* exacerbated abuses by soldiers on treatment blocks during the intervention, though the effects are substantively small and only weakly statistically significant. Importantly, these abuses were exceedingly rare: just 10 out of 2,085 monitoring survey respondents (0.48% of the sample) reported either verbal or physical abuse by soldiers. Equally importantly, all monitoring survey respondents who reported an abuse were surveyed *after* the intervention was complete. (One was surveyed on November 29, 11 days after the end of the intervention; one was surveyed on November 30, one on December 2, six on December 9, and one on December 10.) While we allow for telescoping, it is unclear if these abuses were related to *Plan Fortaleza*.

We find more robust evidence of increased abuses by the police. Relative to control blocks, monitoring survey respondents on treatment blocks were 0.037 percentage points more likely to report an abuse by a police officer. Police abuse was also more common than military abuse, with 72 monitoring survey respondents (3.45% of the sample) reporting abuse by a police officer. Roughly half of the monitoring survey respondents who reported police abuse were surveyed while the intervention was ongoing, and roughly half after. We also find some suggestive evidence that this increase in police abuse persisted over time: residents of spillover blocks were 0.03 percentage points more likely to report an abuse by police officers in the endline survey. We find no evidence of increased police abuse on treatment blocks in the endline survey, and no evidence of increased military abuse on either treatment or spillover blocks. As in the monitoring survey, police abuse was much more common than military abuse in the endline survey: only 99 of our 7,913 endline survey respondents (1.25% of the sample) reported abuse by soldiers, compared to 1,045 (13.21% of the sample) who reported abuse by police officers.

Importantly, however, these adverse effects on human rights abuses are specific to the surveys: we find little to no evidence of increased abuses by either soldiers or police officers in administrative data from the Office of the Attorney-General, nor in the firsthand observations of the civilian monitors. The civilian monitors recorded just one incident of verbal abuse throughout the duration of the study, and no incidents of physical abuse. The incident of verbal abuse occurred on October 16, 2019 at 8:31pm as soldiers were departing a treatment block, apparently at a moment of heightened insecurity.[12] Subsequent interviews with the monitor who observed this incident suggest that it was not serious.

Moreover, as we show in Appendix A.8, only one alleged abuse in the Attorney-General's data occurred in either of the two *comunas* in our sample during the intervention. Two more occurred near but outside of *comuna* 20, and were sufficiently far away from the nearest treatment block—217 meters and 257 meters, respectively—that we believe they were likely unrelated to *Plan Fortaleza*. The one alleged abuse from inside the sample occurred on October 29, at the intersection of two treatment blocks and a spillover block in *comuna* 18. The incident involved a transit police officer who was accused of unfairly restricting a citizen's freedom of movement, likely in relation to existing traffic laws in Cali. Given the nature of the allegation and the individuals involved, we believe it was likely unrelated to *Plan Fortaleza*.

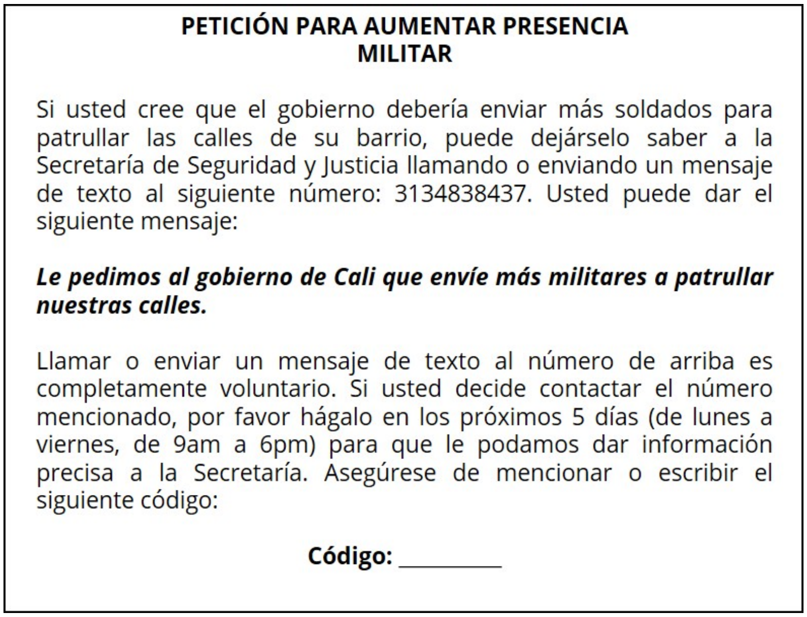
4.3 Political attitudes and behaviors

Positive effects on perceived legitimacy of the military; mixed effects on perceived legitimacy of the police

In Table 3 we report the ITT of the program on the perceived legitimacy of the police and military in the eyes of civilians, and on citizens' demand for military policing. To construct the dependent variable in columns 1 and 2, respondents were asked if they believe the police and military investigate crimes effectively and professionally; use excessive force; are corrupt; treat rich and poor

[12] In a comment accompanying the report, the monitor explained that "for our security, the commanding officer told us to leave quickly."

Figure 2: **Flyer for behavioral measure of demand for military policing**



PETICIÓN PARA AUMENTAR PRESENCIA MILITAR

Si usted cree que el gobierno debería enviar más soldados para patrullar las calles de su barrio, puede dejárselo saber a la Secretaría de Seguridad y Justicia llamando o enviando un mensaje de texto al siguiente número: 3134838437. Usted puede dar el siguiente mensaje:

Le pedimos al gobierno de Cali que envíe más militares a patrullar nuestras calles.

Llamar o enviar un mensaje de texto al número de arriba es completamente voluntario. Si usted decide contactar el número mencionado, por favor hágalo en los próximos 5 días (de lunes a viernes, de 9am a 6pm) para que le podamos dar información precisa a la Secretaría. Asegúrese de mencionar o escribir el siguiente código:

Código: _________

Colombians equally; treat Afro-Colombians and non-Afro-Colombians equally; treat the young and old equally; and treat men and women equally. Responses were recorded on five-point Likert scales, then aggregated into standardized additive indices.

We use our costly behavioral measure of demand for military policing to construct the dependent variable in column 3. At the end of the survey, respondents were given a flyer (depicted in Figure 2) with the phone number of a hotline they could text or call to request additional military patrols in their neighborhood. Each flyer included a unique ID number, allowing us to record the number of texts and calls originating from each block in the sample. Respondents were informed that the petition was voluntary and anonymous, and that there was no guarantee that the government would respond with more military patrols.

Consistent with hypothesis 4, we find that *Plan Fortaleza* improved perceptions of the military in our endline survey by 0.079 standard deviations on treatment blocks (column 2), though this effect is only weakly statistically significant at conventional levels. We find no evidence that the program improved perceptions of the military on spillover blocks. Also consistent with hypothesis 4, we find that the program increased demand for military policing using our costly behavioral

Table 3: **Treatment effects on perceptions of the police and military and demand for military policing**

	Survey data		Behavioral data
	Perceptions of police	**Perceptions of military**	**Demand for military**
Treatment	0.030	0.079*	0.053**
	(0.043)	(0.044)	(0.023)
Spillover	-0.021	-0.002	0.033**
	(0.034)	(0.036)	(0.016)
Individual-level controls	✓	✓	✗
Block-level controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Observations	7794	7757	1167
R^2	0.05	0.04	0.11
Control mean	-0.001	-0.014	0.028

Notes: ITT on standardized additive indices of perceptions of police (column 1) and military (column 2) in the survey and demand for military policing (column 3) in our behavioral measure. Columns 1 and 2 include neighborhood fixed effects and individual- and block-level controls. Column 3 includes neighborhood fixed effects and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors, clustered by block in columns 1 and 2, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

measure (column 3). Compared to control blocks, the hotline received 0.053 more calls and texts from treatment blocks and 0.033 more from spillover blocks. These constitute substantively large increases of 189.3% and 117.9%, respectively, over the control group mean (0.028). Contrary to hypothesis 5, however, we find no evidence that the program diminished perceptions of the police (column 1).

As an additional proxy for perceived legitimacy, the endline also included a conjoint survey experiment inspired by Flores-Macías and Zarkin (2022). Respondents were presented with three pairs of images sequentially. Each image depicted either a soldier or a police officer, carrying either a rifle or a pistol, as in Figure 3.[13] We varied the weapon to test whether respondents' perceptions were driven by a preference for the military specifically (in which case they should have preferred the soldier over the police officer regardless of weapon), militarization more generally (in which case they should have preferred the rifle over the pistol regardless of whether it was carried by a soldier or police officer), or both. For each pair of images, respondents were asked which of the two individuals depicted would make them feel safer, and which was more likely to deter crime, to commit abuses, and to be corrupt. We aggregate responses to these questions into a standardized additive index.

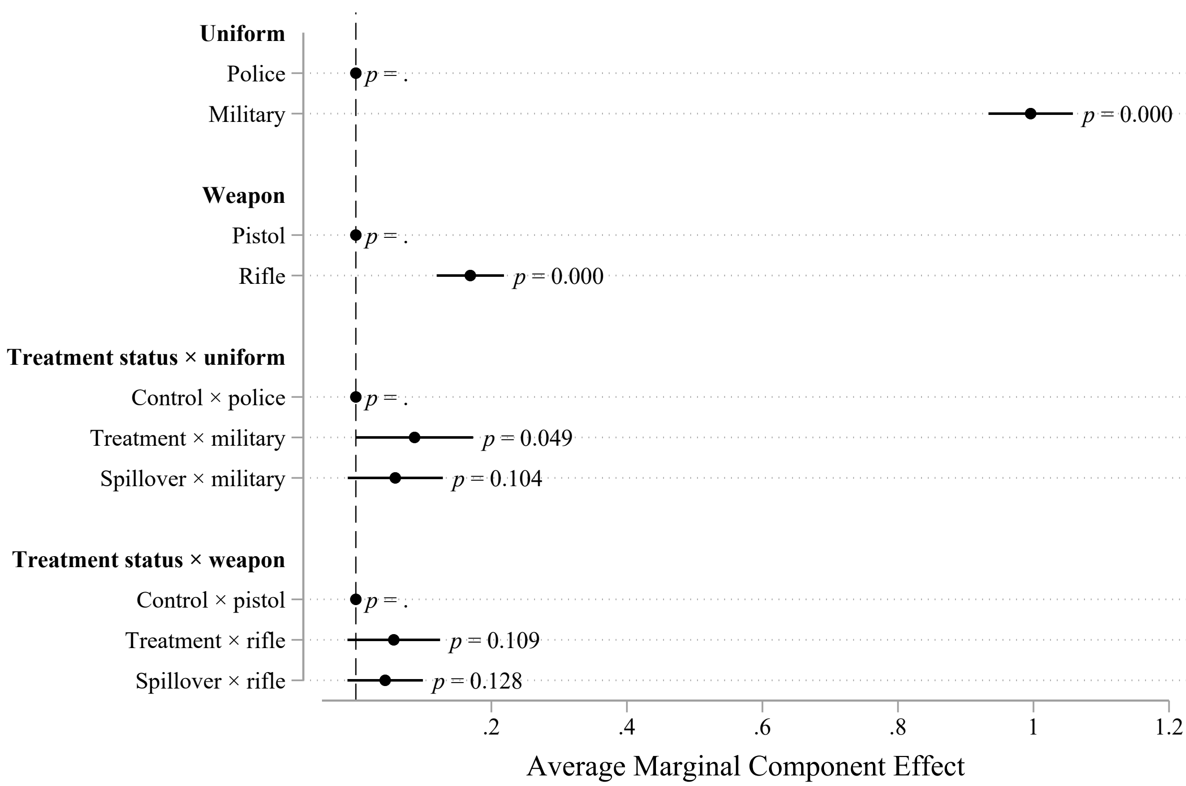
Our use of images rather than text here is advantageous because it more closely approximates the situations in which citizens encounter soldiers and police officers in the real world, especially since many of these encounters are passive, with little interpersonal interaction (Flores-Macías and Zarkin 2022). Images also convey information that respondents might otherwise infer, confounding their responses. For example, if respondents believe soldiers are more likely than police officers to be male (which is true in Colombia), and if they prefer to be policed by men, then a stated preference for soldiers over police officers may be confounded by an unstated preference for men over women. Because the images are identical except for the uniform and weapon, we eliminate many of these potential confounders, including age, gender, race, and physical stature.

Figure 4 plots the Average Marginal Component Effect (AMCE) of each attribute in the

[13] Randomization was restricted such that no respondent saw the same pair more than once. To avoid ordering effects, the position of the images on the screen (left or right) was randomized as well.

Figure 3: **Conjoint survey experiment images**

Figure 4: **Treatment effects on perceptions of the police and military using conjoint experiment**



Notes: ITT on standardized additive indices of perceptions of the police and military in conjoint experiment. All specifications include neighborhood fixed effects and individual- and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. Lines denote 95% confidence intervals.

Table 4: **Treatment effects on support for vigilantism and military coups**

	Survey data			
	Support for military coups		Support for vigilantism	
	In response to crime	In response to corruption	Bypassing the legal system	Bypassing the police
Treatment	0.048**	0.014	0.086**	-0.014
	(0.020)	(0.020)	(0.040)	(0.043)
Spillover	0.048***	0.007	0.061*	0.006
	(0.017)	(0.017)	(0.035)	(0.035)
Individual-level controls	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓
Observations	7806	7830	7882	7868
R^2	0.04	0.04	0.03	0.03
Control mean	0.474	0.664	-0.032	0.020

Notes: ITT on dummies indicating support for military coups in response to crime (column 1) and corruption (column 2), and on indices of support for bypassing the legal system (column 3) and the police (column 4) to punish criminals based on survey data. All specifications include neighborhood fixed effects and individual- and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors, clustered by block, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

conjoint (Hainmueller, Hopkins and Yamamoto 2014), as well as the coefficient on the interaction term between each attribute and the respondent's *Plan Fortaleza* treatment assignment. We find that respondents expressed a strong preference for military policing overall. This preference is driven by positive perceptions of the military specifically: while respondents preferred the rifle over the pistol, they preferred soldiers over police officers by much larger margins. Consistent with hypotheses 4 and 5, we also find that this preference for the military over the police was statistically significantly stronger on treatment blocks relative to control blocks. Respondents' preference for the military was stronger on spillover blocks as well, and their preference for the rifle over the pistol was stronger on both treatment and spillover blocks, though these effects are not quite statistically significant at conventional levels.

Adverse effects on commitment to democracy and the rule of law

In Table 4 we report the ITT of the program on civilians' commitment to democracy and the rule of law. The dependent variables in columns 1 and 2 are dummies indicating whether endline survey respondents agreed that military coups are sometimes justified in response to rampant crime (column 1) or pervasive corruption (column 2). To construct the dependent variable in column 3, respondents were asked on a five-point Likert scale whether they believe it is acceptable to bypass the legal system in order to "immediately address" crimes. For column 4, respondents were asked on a three-point Likert scale whether they believe it is justifiable to beat a criminal rather than deliver him to the police. These questions were taken from the Latin American Public Opinion Project (LAPOP) survey. Both scales are standardized in Table 4 for ease of interpretation.

Consistent with hypothesis 6, we find that the intervention weakened residents' commitment to democracy and the rule of law, though the effects vary by outcome in largely intuitive ways. Residents of both treatment and spillover blocks were 4.8 percentage points more likely to support military coups in response to rampant crime, but not in response to pervasive corruption. This is perhaps unsurprising, as *Plan Fortaleza* aimed first and foremost to deter criminals. Likewise, residents of both treatment and spillover blocks were more likely to support vigilantism to bypass the legal system but not the police. This is consistent with our finding in Appendix A.11.3 that residents became more willing to report crimes and otherwise cooperate with the police (and the military) after the intervention was complete.

Weak or null effects on support for right-wing causes and candidates

In Table 5 we report the ITT on support for right-wing causes and candidates. These analyses exploit two events that occurred during and immediately after the intervention. First, on October 27, 2019, roughly one month after *Plan Fortaleza* began, Colombia held its first municipal elections since the 2016 peace accord. The dependent variable in column 1 is a dummy indicating whether

Table 5: **Treatment effects on turnout for right-wing candidates and support for and participation in left-wing protests**

	Survey data		
	Turnout for conservative candidates	**Attitudes and behavior towards protests**	
	All	Approves	Participated
Treatment	-0.012	-0.077*	-0.005
	(0.010)	(0.042)	(0.011)
Spillover	0.004	-0.008	-0.007
	(0.009)	(0.036)	(0.010)
Individual-level controls	✓	✓	✓
Block-level controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Observations	7908	7419	7498
R^2	0.02	0.09	0.07
Control mean	0.077	-0.001	0.094

Notes: ITT on turnout for right-wing candidates (column 1) and support for (column 3) and participation in (column 4) left-wing protests based on survey data. All specifications include neighborhood fixed effects and individual- and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors, clustered by block, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

respondents voted for right-wing mayoral candidates in Cali's elections.[14] Second, beginning on November 21, 2019, massive nationwide protests erupted throughout Colombia. The protests were associated with left-wing opponents of President Iván Duque's right-wing government. To construct the dependent variable in column 2, respondents were asked whether they approved of the protests on a five-point Likert scale; this scale is standardized for ease of interpretation. For column 3, we code a dummy indicating whether respondents reported participating in the protests.

Contrary to hypothesis 7, we find no evidence that the program increased turnout for right-wing candidates. As we discuss below, this may reflect a peculiarity of Cali's 2019 municipal elections, in which both right- and left-wing candidates were perceived as equally tough on crime and equally supportive of military policing. We do, however, find that respondents on treatment blocks were (weakly) less supportive of left-wing protests (column 2), though they were no less likely to participate in them (column 3). *Plan Fortaleza* reduced support for the protests by 0.077 standard deviations on treatment blocks. We find no evidence that the program reduced support or participation in the protests on spillover blocks.

5 Discussion

Our results suggest that intensive, recurring military patrols targeting crime hot spots in Cali did not reduce the objective prevalence of crime, improve civilians' subjective perceptions of safety, or mitigate human rights abuses. Yet our results also suggest that military patrols increased the perceived legitimacy of the military in the eyes of civilians, generated demand for additional military involvement in domestic law enforcement, and—more troublingly—weakened citizens' commitment to democracy and the rule of law by increasing their support for vigilantism and military coups.

While surprising, this apparently contradictory combination of results is consistent with Flores-Macías and Zarkin (2022, 1390), who find that citizens living in areas where the Mexican

[14] We identified right-wing candidates based on their official campaign records and press releases. To avoid conditioning on a post-treatment variable, we code this dummy as a 0 for respondents who did not vote. Our results are substantively similar if we condition on voting.

military conducted domestic policing operations are more likely to view the military as effective and respectful of civil liberties and more likely to support military policing in their own neighborhoods, despite evidence that these operations exacerbated violence and precipitated human rights abuses (Flores-Macías and Zarkin 2021*a*). Our findings, combined with those of Flores-Macías and Zarkin (2022), pose a puzzle that the literature on penal populism has yet to address: if military policing and other *mano dura* policies are ineffective or even counterproductive, why do citizens continue to demand more of them?

To explore potential answers to this question, we conducted qualitative interviews with 49 civil society leaders from the two *comunas* in our sample. All interviews occurred between June 26 and July 14, 2020, roughly seven months after the end of the intervention. Unfortunately, due to the COVID-19 pandemic, we were unable to return to these *comunas* to interview residents in person. We instead limited our data collection to civil society leaders who could be reached by phone. We confirmed the *comuna* and neighborhood in which respondents lived and worked, but to maintain anonymity we did not ask for their addresses, and so do not know whether their blocks were assigned to treatment, spillover, or control. Nonetheless, these qualitative data help us understand and contextualize our quantitative results.

Our interviews corroborate the demand for military policing that we observe in the endline survey: of the 49 civil society leaders we interviewed, only two opposed military patrols, and these two *also* opposed patrols by the police.[15] Nearly all other respondents voiced strong, unequivocal support for military policing; as one argued, "there's no security here. I'd like the military to get involved and make people obey the law."[16] Even the few who expressed some ambivalence nonetheless described military patrols as necessary to prevent crime.[17] As one of these respondents explained, "do I like it [military policing]? No. But yes, I think it's necessary."[18]

Our interviews also suggest at least two possible reasons why citizens who are exposed to in-

[15] ID 254, July 6, 2020, *comuna* 18, Sector Los Mandarinos Alto Jordán; ID 359, July 14, 2020, *comuna* 18, Desconocido.
[16] e.g. ID 168, June 26, 2020, *comuna* 18, Belén.
[17] e.g. ID 326, July 1, 2020, *comuna* 18, Francisco Eladio.
[18] ID 355, June 30, 2020, *comuna* 20, El Cortijo.

effective or counterproductive *mano dura* policies might nonetheless demand more of them. First, reducing crime may be less important to civilians than mitigating police corruption and abuse, especially if they have witnessed police misconduct firsthand. In our quantitative data we find that *Plan Fortaleza* strengthened Cali residents' preference for military policing while also exacerbating human rights abuses by the police. It is possible that these two results are related—that residents of treatment and spillover blocks were more likely to prefer military policing precisely because they were more likely to observe abuses by police officers, and therefore to be fearful of police presence. Our qualitative data are consistent with this explanation. As one civil society leader observed, "there are many citizens who are more afraid of the police than of criminals themselves."[19] Another explained, "of course, yes, I support them [military patrols]. I live in fear of the police."[20] A third said simply, "what I'm very terrified of is the police."[21]

A second possible explanation is that civilians are more likely to believe (perhaps incorrectly) that military policing reduces crime if they have seen soldiers patrolling firsthand. According to Pion-Berlin and Carreras (2017, 18), support for military policing "is not an endorsement of repression so much as it is a belief—accurate or not—that the military can, compared to the police, more capably combat criminal elements without harming innocent civilians to the same degree." Citizens who have witnessed soldiers patrolling may be especially likely to subscribe to this belief. Consistent with this interpretation, civil society leaders who had observed military patrols in their neighborhoods almost universally believed they were effective. As one respondent explained, "in my neighborhood they did patrols and yes, they helped with security."[22] More tellingly, some seemed to believe that the perception of effectiveness was at least as important as the reality: "the simple fact of seeing their [soldiers'] presence gives people living here a sense of security."[23] (Note, however, that we find little to no evidence of this heightened sense of security in our endline survey.)

[19] ID 159, June 26, 2020, *comuna* 18, Los Chorros.
[20] ID 168, June 26, 2020, *comuna* 18, Belén.
[21] ID 168, June 26, 2020, *comuna* 18, Belén.
[22] e.g. ID 049, June 27, 2020, *comuna* 20, Venezuela; ID 088, July 1, 2020, *comuna* 18, Alto Melendez.
[23] ID 058, July 2, 2020, *comuna* 18, Altos de la Luisa.

These civil society leaders generally attributed the efficacy of military policing to the military's superior logistical and coercive capacity; to the integrity of the armed forces; and, most important, to the respect that soldiers elicit from both civilians and criminals. As one respondent explained, "here in Cali, young people have lots of respect for soldiers. The police they don't respect. The hooligans on the street listen to the military."[24] When criminals encounter the police, "they throw rocks at them, they steal their weapons;"[25] but "do you think they would dare tangle with a soldier?"[26] Even respondents who were aware of the military's deficiencies nonetheless tended to describe soldiers as more effective than police officers. As one respondent observed, "nobody respects the police because of all the corruption they've been involved in. It's not that there's no corruption in the military; obviously there is. But soldiers are less likely to be corrupt, and they get much more respect when they walk down the street. That's something that's missing in these neighborhoods: respect for authority."[27]

Interestingly, while we find that *Plan Fortaleza* generated demand for additional military policing, we find no evidence that it increased support for the right-wing parties and politicians with whom military policing is usually associated. Our qualitative interviews help resolve this apparent paradox in the quantitative data. In particular, the interviews reveal that the partisan divide on military policing is not especially salient in Cali. When asked about the 2019 municipal elections, most of the civil society leaders we interviewed were unfamiliar with the candidates' platforms, and were unsure which ones supported military policing. 15 respondents—a plurality—simply did not know the candidates' positions on the issue. 13 respondents associated military policing with left-wing candidates, and 12 associated it with right-wing candidates. Four associated military policing with *all* candidates. As one respondent explained, "all of the candidates supported them [military patrols];"[28] another claimed that "all the candidates" and "all good people and the whole of the city [supported the patrols]... It's not about one side or another."[29]

[24]ID 234, June 27, 2020, *comuna* 20, El Cortijo.
[25]ID 312, June 27, 2020, *comuna* 18, Caldas.
[26]ID 018, July 2, 2020, *comuna* 18, Alto Melendez.
[27]ID 358, July 14, 2020, *comuna* 18, Desconocido.
[28]ID 122, June 26, 2020, *comuna* 18, Lourdes
[29]ID 071, June 27, 2020, *comuna* 18, Polvorines.

A detailed desk review of the platforms of the top four mayoral candidates suggests that their positions on this issue were, indeed, indistinguishable. Only one of the top four candidates—Danis Rentería, who served more than 24 years in the army—included military involvement in domestic policing operations as an explicit component of his platform.[30] The platforms of the other top three candidates did not mention military policing. This helps explain why most respondents either did not know which candidates supported military policing or believed that all candidates supported it equally. It also helps explain why we find no evidence of increased turnout for right-wing candidates in response to *Plan Fortaleza*.

6 Conclusion

Governments throughout Latin America and other parts of the Global South increasingly rely on their armed forces to engage in domestic policing operations. Military policing is motivated in part by concerns about corruption, abuse, and incompetence among Latin American police forces. Yet for all the popularity of military policing, empirical evidence regarding its efficacy remains limited. We help fill this gap with what we believe to be the first experimental evaluation of this increasingly common *mano dura* approach to law enforcement.

Using a combination of administrative crime and human rights data, surveys, a conjoint survey experiment, a costly behavioral measure, qualitative interviews, and firsthand observations by civilian monitors, we find little to no evidence that military policing reduced the objective prevalence of crime, improved civilians' subjective perceptions of safety, or mitigated human rights abuses in Cali, one of the world's most violent cities. Yet despite these null or even adverse effects on security and human rights, we also find that military policing strengthened demand for more aggressive military involvement in law enforcement and other aspects of governance. We use our qualitative interviews to help reconcile this puzzling combination of results.

These findings have important theoretical implications for our understanding of the in-

[30]Specifically, Rentería proposed "integrating the military with the police." See https://www.elpais.com.co/elecciones-2019/conozca-las-propuestas-de-danis-renteria-para-llegar-a-la-alcaldia-de-cali.html.

creasingly blurred distinction between the military and the police in ostensibly democratic countries. This distinction has long been considered a "preeminent feature of the modern nation-state" (Kraska 2007, 501; see also Giddens 1987; Huntington 1957), and some scholars have expressed concern about the macro-level repercussions of military policing for democracy, the rule of law, and civil-military relations (Dunlap Jr. 1999; Ricks 1997). Our results suggest that the micro-level consequences may be equally profound, reshaping citizens' understanding not just of law enforcement, but of governance more generally. Military policing is usually framed as a necessary temporary measure to combat crime. But fusing the roles of the military and the police in the short term may generate increased demand for greater fusion in the long run, even when soldiers prove no more effective than police officers at deterring criminals.

From a more practical perspective, our results suggest that military policing can generate adverse unintended consequences that policymakers may (and should) prefer to avoid. Advocates claim that military policing deters crime and mitigates abuses by predatory police forces. Our results lend little to no credence to these claims; if anything, the opposite is true. Moreover, and possibly more important, our results suggest that military policing may shape citizens' attitudes and behaviors in ways that are detrimental to democracy and the rule of law—another adverse unintended consequence of military policing that is rarely contemplated by its proponents. Policymakers should consider alternative crime reduction strategies that are less likely to produce perverse effects.

As with any study focused on a single site, we cannot be sure how far our results will generalize. As noted above, however, Cali is an important test case in and of itself, and it shares a number of salient characteristics with other cities where military policing is currently being implemented, including extreme socioeconomic inequality; a concentration of crime in the city's most disadvantaged neighborhoods; and a police force plagued by accusations of corruption, abuse, and collusion with organized crime. While Cali suffers from disproportionately high rates of violent crime, this is true of most cities in which military policing is currently being implemented or considered. (Indeed, it is precisely these high violent crime rates that make military policing seem necessary in the first place.)

Moreover, while the program we evaluate was short, and while it targeted only a subset of the city's neighborhoods, it was similar to many military policing interventions in its temporal and geographical scope (Flores-Macías and Zarkin 2021*a*, 526). Our evaluation is also analogous to studies of hot spots policing, which similarly focus on "place-based" interventions involving a concentration of resources on specific neighborhoods within particular cities for a limited amount of time (Blattman et al. 2021; Braga et al. 2019; Collazos et al. 2019). While we can only speculate as to the generalizability of our results to settings that are very different from Cali, our findings, combined with results from other recent studies (Espinosa and Rubin 2015; Flores-Macías 2018; Flores-Macías and Zarkin 2021*a*; Flores-Macías and Zarkin 2021*b*; Osorio 2015), provide reason for caution as policymakers consider similar *mano dura* interventions in Colombia and beyond.

References

Abadie, Alberto, Susan Athey, Guido W Imbens and Jeffrey M Wooldridge. 2023. "When Should You Adjust Standard Errors for Clustering?" *The Quarterly Journal of Economics* 138(1):1–35.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103(484):1481–1495.

Aronow, Peter M. and Cyrus Samii. 2017. "Estimating Average Causal Effects under General Interference, with Application to a Social Network Experiment." *The Annals of Applied Statistics* 11(4):1912–1947.

Bayley, David H. 1998. "Criminal Investigation: Introduction." *What Works in Policing* pp. 71–74.

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, Donald P. 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.

Bove, Vincenzo and Evelina Gavrilova. 2017. "Police Officer on the Frontline or a Soldier? The Effect of Police Militarization on Crime." *American Economic Journal: Economic Policy* 9(3):1–18.

Braga, Anthony A., Brandon S. Turchan, Andrew V. Papachristos and David M. Hureau. 2019. "Hot spots policing and crime reduction: an update of an ongoing systematic review and meta-analysis." *Journal of experimental criminology* 15(3):289–311.

Braga, Anthony A. and Brenda J. Bond. 2008. "Policing Crime and Disorder Hot Spots: A Randomized Controlled Trial." *Criminology* 46(3):577–607.

Brinks, Daniel M. 2007. *The Judicial Response to Police Killings in Latin America: Inequality and the Rule of Law*. Cambridge, UK: Cambridge University Press.

Campbell, Donald J. and Kathleen M. Campbell. 2010. "Soldiers as Police Officers/Police Officers as Soldiers: Role Evolution and Revolution in the United States." *Armed Forces & Society* 36(2):327–350.

Collazos, Daniela, Eduardo García, Daniel Mejía, Daniel Ortega and Santiago Tobón. 2019. "Hot Spots Policing in a High Crime Environment: An Experimental Evaluation in Medellín." SSRN Scholarly Paper 3316968.

Dammert, Lucia and John Bailey. 2007. "¿Militarización de La Seguridad Pública En América Latina?" *Foreign Affairs En Español* April-June:61–70.

De Bruin, Erica. 2022. "Policing Insurgency: Are More Militarized Police More Effective?" *Small Wars & Insurgencies* 33(4-5):742–766.

Delehanty, Casey, Jack Mewhirter, Ryan Welch and Jason Wilks. 2017. "Militarization and police violence: The case of the 1033 program." *Research & politics* 4(2):1–7.

Dunlap Jr., Charles J. 1999. "The Police-Ization of the Military." *Journal of Political and Military Sociology* 27:397–418.

Espinosa, Valeria and Donald B. Rubin. 2015. "Did the Military Interventions in the Mexican Drug War Increase Violence?" *The American Statistician* 69(1):17–27.

Flores-Macías, Gustavo A. 2018. "The Consequences of Militarizing Anti-Drug Efforts for State Capacity in Latin America: Evidence from Mexico." *Comparative Politics* 51(1):1–20.

Flores-Macías, Gustavo A. and Jessica Zarkin. 2021*a*. "The Militarization of Law Enforcement: Evidence from Latin America." *Perspectives on Politics* 19(2):519–538.

Flores-Macías, Gustavo and Jessica Zarkin. 2021*b*. "The Consequences of Militarized Policing for Human Rights: Evidence from Mexico." Presented at the International Studies Association Annual Conference, April 6–9.

Flores-Macías, Gustavo and Jessica Zarkin. 2022. "Militarization and Perceptions of Law Enforcement in the Developing World: Evidence from a Conjoint Experiment in Mexico." *British Journal of Political Science* 52(3):1377–1397.

Gelman, Andrew. 2003. "A Bayesian Formulation of Exploratory Data Analysis and Goodness-of-Fit Testing." *International Statistical Review / Revue Internationale de Statistique* 71(2):369–382.

Giddens, Anthony. 1987. *The Nation-State and Violence*. Berkeley: University of California Press.

González, Yanilda María. 2020. *Authoritarian Police in Democracy: Contested Security in Latin America*. Cambridge, UK: Cambridge University Press.

Greener-Barcham, B.K. 2007. "Crossing the Green or Blue Line? Exploring the Military–Police Divide." *Small Wars & Insurgencies* 18(1):90–112.

Gunderson, Anna, Elisha Cohen, Kaylyn Jackson Schiff, Tom S. Clark, Adam N. Glynn and Michael Leo Owens. 2021. "Counterevidence of Crime-Reduction Effects from Federal Grants of Military Equipment to Local Police." *Nature Human Behaviour* 5(2):194–204.

Hainmueller, Jens, Daniel J. Hopkins and Teppei Yamamoto. 2014. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis* 22(01):1–30.

Harris, Matthew C., Jinseong Park, Donald J. Bruce and Matthew N. Murray. 2017. "Peacekeeping Force: Effects of Providing Tactical Equipment to Local Law Enforcement." *American Economic Journal: Economic Policy* 9(3):291–313.

Holland, Alisha C. 2013. "Right on Crime?: Conservative Party Politics and Mano Dura Policies in El Salvador." *Latin American Research Review* 48(1):44–67.

Huntington, Samuel P. 1957. *The Soldier and the State: The Theory and Politics of Civil–Military Relations*. Cambridge, MA: Belknap.

Keele, Luke. 2015. "The Statistics of Causal Inference: A View from Political Methodology." *Political Analysis* 23(3):313–335.

Kraska, Peter B. 2007. "Militarization and Policing–Its Relevance to 21st Century Police." *Policing* 1(4):501–513.

Kraska, Peter B and Victor E Kappeler. 1997. "Militarizing American Police: The Rise and Normalization of Paramilitary Units." *Social Problems* 44(1):1–18.

Lessing, Benjamin. 2015. "Logics of Violence in Criminal War." *Journal of Conflict Resolution* 59(8):1486–1516.

Lowande, Kenneth. 2021. "Police Demilitarization and Violent Crime." *Nature Human Behaviour* 5(2):205–211.

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.

Magaloni, Beatriz and Luis Rodriguez. 2020. "Institutionalized Police Brutality: Torture, the Militarization of Security, and the Reform of Inquisitorial Criminal Justice in Mexico." *American Political Science Review* 114(4):1013–1034.

Muggah, Robert, Juan Carlos Garzón and Manuela Suárez. 2018. *Mano Dura: The costs and benefits of repressive criminal justice for young people in Latin America*. Rio de Janeiro: Igarapé Institute.

Mummolo, Jonathan. 2018. “Militarization Fails to Enhance Police Safety or Reduce Crime but May Harm Police Reputation.” *Proceedings of the National Academy of Sciences* 115(37):9181–9186.

Osorio, Javier. 2015. “The Contagion of Drug Violence: Spatiotemporal Dynamics of the Mexican War on Drugs.” *Journal of Conflict Resolution* 59(8):1403–1432.

Pion-Berlin, David and Miguel Carreras. 2017. “Armed Forces, Police and Crime-Fighting in Latin America.” *Journal of Politics in Latin America* 9(3):3–26.

Pratt, John. 2007. *Penal Populism*. London: Routledge.

Ricks, Thomas E. 1997. “The Widening Gap Between Military and Society.” *The Atlantic* July.

Rosenbaum, Dennis P. 2006. The Limits of Hot Spots Policing. In *Police Innovation: Contrasting Perspectives*, ed. David Weisburd and Anthony A. Braga. Cambridge, UK: Cambridge University Press pp. 245–266.

Tecott, Rachel and Sara Plana. 2016. “Maybe U.S. Police Aren’t Militarized Enough. Here’s What Police Can Learn from Soldiers.” *Monkey Cage* August 16.

Vanegas Muñoz, Gildardo. 2021. *La saga del narcotráfico en Cali, 1950-2018.* Programa Editorial UNIVALLE.

Visconti, Giancarlo. 2020. “Policy Preferences after Crime Victimization: Panel and Survey Evidence from Latin America.” *British Journal of Political Science* 50(4):1481–1495.

Wang, Ye, Cyrus Samii, Haoge Chang and P.M. Aronow. 2022. “Design-Based Inference for Spatial Experiments with Interference.” arXiv working paper.
URL: *https://arxiv.org/abs/2010.13599*

Weisburd, David and Cody W. Telep. 2014. “Hot Spots Policing: What We Know and What We Need to Know.” *Journal of Contemporary Criminal Justice* 30(2):200–220.

Wood, Nathan A. 2015. “The Ferguson Consensus Is Wrong: What Counterinsurgency in Iraq and Afghanistan Teaches Us about Police Militarization and Community Policing.” *Lawfare Research Paper Series* 3(1):1–22.

Zechmeister, Elizabeth J., ed. 2014. *The Political Culture of Democracy in the Americas, 2014: Democratic Governance across 10 Years of the AmericasBarometer*. Washington, DC: USAID.

APPENDIX

A.1 POWER CALCULATIONS

We estimate minimum detectable effects (MDEs) using administrative crime data from before the start of the intervention. (For simplicity, we focus on crime alone, as this was the only outcome for which we had pre-treatment data before registering our pre-analysis plan. We likely have higher power to detect effects on other outcomes, for reasons discussed below.) We first replicate our randomization procedure 1,000 times. Within each replication we simulate the treatment and spillover effects of the intervention, knowing that the true effect is 0. The dependent variable is identical to the one in columns 1 and 2 of Table 1 in the paper, except that it is standardized for ease of interpretation, and only includes crimes committed during the four-month period between July 1 and September 29, 2019—the day before the intervention began. All replications include neighborhood fixed effects, inverse probability weights, and a lagged dependent variable for the number of crimes committed on each block between January 1 and June 30, 2019.

Formally, we estimate our MDEs using a weighted least squares regression given by

$$y_{jk} = \theta t_{jk} + \lambda s_{jk} + \gamma y_{jk,t-1} + \alpha_k + \epsilon_{jk} \qquad \text{(A.1)}$$

where y_{jk} denotes crime on block j in neighborhood k; t_{jk} denotes assignment to treatment; s_{jk} denotes assignment to spillover; $y_{jk,t-1}$ denotes the lagged dependent variable; α_k denotes neighborhood fixed effects; and ϵ_{jk} is a block-level error term. Observations were weighted by the inverse probability of assignment to their realized treatment status in each replication. We then calculate the standard deviation of these simulated treatment and spillover effects. To estimate our MDEs, we simply multiply the standard deviation of the simulated effects by 2.49.

Without covariates we find that we are powered to detect treatment effects of approximately 0.18 standard deviations, and spillover effects of approximately 0.14 standard deviations.[1] These

[1]Our MDE estimates are somewhat sensitive to the temporal window we use to define the dependent variable: they

would generally be considered small MDEs, meaning that we should be able to detect even substantively modest reductions in crime. These MDEs are also likely conservative, since they do not include covariates, and since the dependent variable is operationalized at the block level. Our inclusion of covariates and our use of survey data (which increases our sample size) for most analyses in the paper should further improve our statistical power.

A.2 Randomization

To classify spillover blocks, we first geocoded all points on all blocks in the sample. We then identified all blocks with at least one point that fell within 25 meters of at least one point on another block. We defined these as adjacent blocks. If one block was assigned to treatment, then all adjacent blocks that were not also assigned to treatment were assigned to spillover. This approach excludes from the spillover group any blocks separated from an adjacent treatment block by a park, highway, or some other barrier.

A.3 Balance tests

Tables A.1 and A.2 report balance tests using our administrative crime data and endline survey data, respectively.

A.4 Implementation

Table A.3 provides descriptive statistics on soldiers' activities while on patrol based on the firsthand accounts of civilian monitors.

become larger as the temporal window expands, and narrower as the temporal window contracts. But they remain small regardless.

A.5 Endline survey

The endline survey sample consists of 1,168 blocks across 30 neighborhoods in *comunas* 18 and 20. The experimental sample consists of 1,254 blocks, but 86 were excluded from the endline survey sample. 70 of these blocks had neither residences nor businesses (e.g. because they were parks); 11 had too few buildings; and 5 were impossible to survey due to security concerns. Seven surveys were administered per block. We surveyed residents and business owners in proportion to the number of households and businesses on each block.

A.6 Ethics

Prior to the intervention we created a comprehensive risk management strategy aimed at identifying potential risks and reducing the probability that they would occur.

Risks to civilians

We identified several potential risks to civilians. First was the risk that our evaluation of *Plan Fortaleza* would divert military attention and resources away from the areas that needed them most. We determined that the probability of this risk materializing was low. *Plan Fortaleza* predated our evaluation, and was designed by municipal authorities in collaboration with the Colombian military. Cali's Security and Justice Secretariat selected the two *comunas* where the program would be implemented; we had no role in this decision. Within these two *comunas*, the Secretariat believed that all neighborhoods and blocks were equally likely to become hot spots for crime, and thus equally likely to benefit from the intervention. We randomized from within this pool of blocks. We also coordinated with the Secretariat and the military to define the length and number of soldiers on each patrol and the number of blocks patrolled each night to ensure the military could continue its other operations in the city without interruption.

Second was the (opposite) risk that military patrols would exacerbate human rights abuses against civilians, as critics of military policing often claim. Assessing the likelihood that this risk would materialize involved weighing the concerns of critics against the perceptions of Colombian elected officials and civilians, who, as discussed in the paper, tend to believe soldiers are less likely than police officers to commit abuses, and who tend to support military policing. Whether military patrols mitigate or exacerbate abuses is an open empirical question to which we believed our study could provide an empirical answer. We believed that answering this question was especially urgent given that *Plan Fortaleza* was scheduled to continue with or without our evaluation of it, and given the prevalence of military policing in Colombia and other Latin American countries.

We also believed, however, that it was our ethical responsibility to report any abuses we observed as quickly as possible. To that end, we collected abuse allegations reported to the Office of the Attorney-General by victims and witnesses, and firsthand observations of abuses from the civilian monitors hired to accompany the soldiers while on patrol. As discussed in the paper, we find no evidence of increased abuses by either the police or the military in either of these data sources. We also used the monitoring survey to track abuses while the intervention was ongoing. Monitoring survey respondents were asked how many times they had seen or heard about physical or verbal abuses committed by police officers or soldiers on their block in the past two weeks. As noted in the paper, military abuse was exceptionally rare: only 10 of 2,085 monitoring survey respondents reported *either* verbal *or* physical abuse by soldiers. None of these respondents was surveyed while the intervention was ongoing. 72 monitoring survey respondents reported at least one abuse by the police, although these reports are clustered geographically—55% of all reports originated in just three neighborhoods (Lleras Camargo, Sector Alto Jordán, and Siloé)—suggesting that multiple respondents may have reported the same incident.

Third was the risk that randomization of military patrols would shift the equilibrium distribution of gang presence and activity in Cali, potentially exacerbating violence. After consulting with the Security and Justice Secretariat and the Colombian military, we determined that the probability of this risk materializing was low. The military already (non-randomly) varied its patrol routes from day to day to prevent criminals from adapting to its presence; this should have prevented an equilibrium distribution of gang presence and activity from emerging. Indeed, neither the military nor the Secretariat believed that such an equilibrium distribution existed in these two *comunas* in the first place. It was highly unlikely that a new, more violent equilibrium would arise as a result of our evaluation, given that no such equilibrium existed before the evaluation began.

Fourth was the risk that civilians would face reprisals for participating in our surveys. To minimize this risk, all surveys were conducted in private before noon each day, and respondents were repeatedly reminded that their participation in the survey was voluntary and anonymous; that the survey could be halted at any time; and that they could skip any question they did not want to answer. Both before and during data collection, we consulted local staff, field supervisors, and civil society representatives to diagnose whether particular blocks posed especially acute security concerns. We adjusted our data collection protocols accordingly. For example, on three blocks (approximately 0.25% of the sample, all of them in the spillover group) we opted not to administer the costly behavioral measure due to safety concerns. Gangs were particularly active on these blocks, and we deemed it too dangerous to leave behind materials that might raise suspicions about citizen collaboration with state security forces.

RISKS TO ENUMERATORS

We identified potential risks to enumerators as well. The Colombian military and the Security and Justice Secretariat selected the two *comunas* in our sample because they were believed to be hot spots for crime. Assuming these assessments were accurate, there was a risk that enumerators would face threats to their safety during implementation of the surveys. We took several precautions to mitigate this risk. As discussed above, we consulted local staff, field supervisors, and civil society representatives to identify potentially problematic blocks. These blocks were surveyed by a team of seven "elite" enumerators with extensive experience surveying vulnerable populations in insecure settings. The team received additional training from field coordinators working in the city of Medellín on a similar survey. The team also followed especially strict security protocols. For example, in one of the more dangerous neighborhoods (Siloé, in *comuna* 20), enumerators were accompanied by six local leaders who were known to community members and who could provide guidance on the most difficult blocks in the neighborhood. These local leaders were enlisted only to help ensure the safety of the enumerators and were not present during survey administration.

RISKS TO MONITORS

Finally, we identified potential risks to the civilian monitors hired to accompany the soldiers while on patrol. First was the risk that criminals would identify and harass or threaten the monitors. To mitigate this risk, we recruited monitors who did not live in the two *comunas* in the study, thus reducing the probability that they would be identified. We also established a direct line of communication with the Security and Justice Secretariat that the monitors could use if they suspected they were being followed or watched. We also developed detailed security protocols in the unlikely event that a monitor was identified. These protocols stipulated, for example, that the Security and Justice Secretariat would provide protection for the monitor for at least two weeks following the incident, and that the monitor would be transferred to the other *comuna* in the sample, or temporarily removed from the project altogether. We received no reports of monitors being followed or threatened at any point during the evaluation.

Second was the risk that monitors might inadvertently get separated from the soldiers while on patrol, or, worse, might get caught in the crossfire of gun battles between soldiers and criminals. While we were (and are) unaware of any such gun battles in Cali prior to the start of the intervention, we took several precautions to mitigate this risk. Monitors were instructed to remain in their vehicle at all times. Monitors also received more advanced training on managing threats to their safety, and were provided with bulletproof vests and armbands clearly identifying them as civilians to minimize the risk that they would be mistaken for soldiers and attacked. We received no reports from monitors that they had become separated from the patrols, nor that the patrols came under fire from criminals.

Third was the risk that soldiers themselves would harass or intimidate the monitors while on patrol, especially if they believed the monitors were reporting on abuses. We again took multiple precautions to minimize this risk. Monitors were trained to file any reports of abuse discretely, using the same smartphone app that they used to help soldiers navigate to the treatment blocks. All soldiers who participated in the intervention were informed that the monitors were working in close partnership with the Security and Justice Secretariat, and that immediate remedial action would be taken against anyone caught harassing them. Finally, we instructed monitors not to disclose any personally identifiable information (beyond their first names) to soldiers. Monitors' last names and ID numbers were made available to high-ranking officials within the Security and Justice Secretariat, but not to the soldiers responsible for conducting patrols.

A.7 Deviations from PAP

In our PAP we proposed to test the ITT of the *Plan Fortaleza* program on both weighted and unweighted indices of crime in the administrative data, with weights corresponding to the average prison sentence associated with each crime under Colombian law. In our PAP we proposed to focus on results using the weighted index; in the paper we instead focus on results using the unweighted index. Based on feedback from criminologists, we determined that our approach to weighting is atypical and yields results that are difficult to interpret. We also proposed to test the ITT of the program on crimes committed during the intervention, disaggregated by day and time. These results are null; we omit them here for compactness. (These and all other omitted results are available upon request.)

In our PAP we also proposed to test the ITT of the program on violent and non-violent crimes separately. We omit these results for compactness. In addition, in our PAP we proposed to collect administrative crime data on homicides, assaults, thefts, car thefts, and motorcycle thefts. We were in fact able to collect administrative crime data on homicides, robberies (including armed robbery and all types of theft), illegal drug sales, and illegal possession of weapons. To avoid discarding potentially useful data, we include all these crimes in our index. We also proposed to test the ITT of the program on arrests based on administrative data. Unfortunately we were unable to obtain these data from the government of Cali, and so we drop this analysis here.

In our PAP we proposed to test the ITT of the program on residents' perceptions of the police and military in the monitoring survey. We dropped most questions about perceptions from the monitoring survey, and so omit these analyses here, instead focusing on perceptions in the endline survey only. In our PAP we also proposed to test the ITT of the program on residents' approval of police and military responses to the massive nationwide strikes that coincided with the end of the intervention. We dropped these questions from the endline survey, and so omit these analyses here as well. We also proposed to test for heterogeneous treatment effects (HTEs) by prior crime rate, prior crime victimization, and gender. These HTEs are almost uniformly null, and we omit them here for compactness.

As noted in the paper, we expand two of the hypotheses from our PAP. First, in our PAP we hypothesized that military policing would increase support for retributive criminal justice policies. In the paper we broaden this hypothesis to encompass commitment to democracy and the rule of law more generally; the ITT on support for retributive justice specifically is null. Second, in our PAP we hypothesized that military policing would increase turnout for right-wing candidates. In the paper we broaden this hypothesis to encompass support for right-wing causes more generally. We report the ITT on turnout for right-wing candidates in column 1 of Table 5.

In our PAP we proposed to compute Lee bounds to test the sensitivity of our results to attrition in the endline survey. Because attrition was so minimal, the Lee bounds are not informative, and we omit them here. Finally, we proposed to conduct complier average causal effect (CACE) analyses to adjust for non-compliance during the intervention. Because compliance was so high, and because it is difficult to define "non-compliance" in the context of recurring military patrols,[2] we omit these analyses here. We also proposed to estimate more complex spillover dynamics using a marginalized individualistic response function, following (Wang et al. 2022). We decided to drop this analysis because the procedure is relatively new and untested.

[2] All treatment blocks were patrolled multiple times, but some were patrolled more frequently than others. It is not obvious what "non-compliance" implies in this context. Likewise, some spillover and control blocks were (likely) patrolled at least once, but this is difficult to ascertain using our data.

A.8 Abuses

Figure A.1 shows the distribution of alleged abuses by state security forces in the two *comunas* in our sample during the intervention, as reported by victims and witnesses to the Office of the Attorney-General.

A.9 Spillover

A.9.1 Testing for decay and saturation

In the paper we assume that spillover effects can only be transmitted from treatment blocks to adjacent control blocks. In Table A.4 we relax this assumption in three ways. First, we assume that treatment effects decay as a linear function of distance to the nearest treatment block. We estimate

$$y_{ijk} = \theta t_{jk} + \lambda \sum_{j=1}^{J} f(d_{jk}) + \beta \mathbf{X_{jk}} + \delta \mathbf{Z_{ijk}} + \alpha_k + \epsilon_{ijk} \tag{A.2}$$

where $f(d_{jk})$ is a linear decay function with a standardized distribution and $f(d_{jk}) = \frac{1}{d_{jk}}$. This function is a weighted sum of distances to all treatment blocks, where *t* indicates treatment blocks. We calculate the distance, *d*, to all treatment blocks; $\frac{1}{d}$ for each block; sum the distances to all treatment blocks for each block; and standardize the resulting variable. The quantity of interest represents the expected increase (or decrease) in crime as a given block is closer by one standard deviation to the nearest treatment block. Because the blocks in our sample are so close to one another, we do not view this specification as especially informative.

Second, we instead assume that treatment effects decay as an exponential function of distance to the nearest treatment block. We estimate

$$y_{ijk} = \theta t_{jk} + \lambda \sum_{j=1}^{J} g(d_{jk}) + \beta \mathbf{X_{jk}} + \delta \mathbf{Z_{ijk}} + \alpha_k + \epsilon_{ijk} \tag{A.3}$$

where $g(d_{jk})$ is a spatial decay function with a standardized distribution and $g(d_{jk}) = \frac{1}{e^{d_{jk}}}$. We calculate the distance, *d*, to all treatment blocks; $\frac{1}{e^{d}}$ for each block; sum the distances to all treatment blocks for each block; and then standardize the resulting variable. The quantity of interest represents the expected increase (or decrease) in crime as a given block is closer by one exponentiated standard deviation to the nearest treatment block. Again, because the blocks in our sample are so close to one another, we do not view this specification as especially informative either.

Finally, we assume that the strength of the treatment effect on any given treatment block is a function of the proportion of adjacent blocks that are also treated. We then reestimate our models using this proportion as the independent variable of interest, restricting our sample to the treatment group only. We estimate

$$y_{jk} = \theta \left(p \times t_{jk}\right) + \beta \mathbf{X_{jk}} + \alpha_k + \epsilon_{jk} \tag{A.4}$$

where y_{jk} denotes the outcome for block j in neighborhood k; p denotes the proportion of adjacent blocks assigned to the treatment group; $\mathbf{X_{jk}}$ denotes block-level covariates; α_k denotes neighborhood fixed effects; and ϵ_j is a block-level error term. All analyses include block-level controls, neighborhood fixed effects, and inverse probability weights.

A.10 Multiple comparison corrections and randomization inference

In Tables A.5 through A.9 we replicate our analyses in the paper using randomization inference (RI) to compute exact p-values and applying multiple comparisons corrections to control the risk of a Type I error. To calculate RI p-values, we first simulate our randomization procedure 1,000 times. Within each simulation, we estimate the ITT on each of our dependent variables and store the coefficients. For each coefficient, we calculate the proportion of simulations in which the absolute value of the simulated coefficient is greater than the absolute value of that same coefficient in the paper. This corresponds to a two-tailed test under the sharp null hypothesis of no treatment or spillover effect for any unit.

In the paper we use indexing to reduce the number of hypotheses we test and thus control the risk of a Type I error. Here we instead apply multiple comparisons corrections within (but not across) "families" of outcome. Table A.5, for example, amounts to a test of the hypothesis that Plan Fortaleza affected security at all, across all of our various proxies for security. Because we generally are not interested in comparing the treatment and spillover effects to one another, we assume each model tests a single combined hypothesis of any effect on the corresponding outcome; our results are substantively similar if we assume each model tests two separate hypotheses. Each table reports the Benjamini-Hochberg q-value and Holm-Bonferroni threshold for our treatment and spillover indicators. Following Anderson (2008), the Benjamini-Hochberg q-value is the smallest false discovery rate at which the null hypothesis will be rejected. The Holm-Bonferroni threshold is the adjusted p-value threshold below which the null hypothesis will be rejected at significance level $\alpha = 0.05$.

A.11 Robustness checks and ancillary analyses

A.11.1 Abuses in monitoring survey disaggregated by date of survey

Table A.10 replicates our results in columns 1 and 2 of Table 2 using alternative coding rules to define the period during and after the intervention. In columns 1, 2, 5, and 6 we assume that any abuses reported by respondents who were surveyed before November 19, 2019 (the day after the intervention ended) occurred during the intervention, and that any abuses reported by respondents who were surveyed after November 19 occurred after the intervention. This may be misleading, however, because respondents were asked to report abuses that they witnessed or heard about in the two weeks prior to the survey. As an additional robustness check, in columns 3, 4, 7, and 8 we assume that any abuses reported by respondents who were surveyed before December 2 (two weeks after the end of the intervention) occurred during the intervention, and that any abuses reported by respondents who were surveyed after December 2 occurred after the intervention.

A.11.2 Police and military activities

Table A.11 reports the ITT of the *Plan Fortaleza* program on measures of police and military activities from the monitoring and endline surveys. We asked monitoring survey respondents how often they had seen or heard about police officers or soldiers visiting their blocks in the prior two weeks (columns 1 and 2), and how often they had seen or heard about police officers or soldiers making arrests on their blocks in the past two weeks (columns 3 and 4). Frequency was measured on a Likert scale from 1 to 4. We asked the same questions at endline, although we extended the temporal window from two weeks to one month (columns 5–8).

A.11.3 Cooperation with authorities

Table A.12 reports the ITT on cooperation with the authorities in the endline survey. To measure cooperation, respondents were asked if they had seen or heard of someone contacting the police or military to alert them to suspicious or criminal activity on the block in the last month, and if they had seen or heard of someone on the block providing information to the police or military to assist with a criminal investigation in the last month. We construct standardized additive indices of cooperation using responses to these questions.

Table A.1: **Balance using administrative data**

	Control	Treatment	Spillover
Panel A: Index without controls			
Weighted standardized crime index	0.001	0.002	-0.003
	(0.013)	(0.019)	(0.010)
Panel B: Index components without controls			
Homicides	-0.019	-0.002	0.021
	(0.057)	(0.066)	(0.042)
Robberies	0.000	0.002	-0.002
	(0.009)	(0.011)	(0.004)
Drug dealing	-0.022	0.073	-0.051
	(0.084)	(0.100)	(0.061)
Attempted homicides	0.153*	-0.108	-0.045
	(0.081)	(0.098)	(0.053)
Illegal possession of a firearm	-0.083	0.022	0.061
	(0.072)	(0.126)	(0.095)
Panel C: Index with controls			
Weighted standardized crime index	0.001	0.006	-0.007
	(0.014)	(0.019)	(0.010)
Number of buildings on block	0.001	-0.000	-0.001
	(0.001)	(0.001)	(0.001)
Area of block	-0.000	-0.000	0.000**
	(0.000)	(0.000)	(0.000)
Distance to nearest army battalion (meters)	0.000**	-0.000*	-0.000
	(0.000)	(0.000)	(0.000)
Distance to nearest police station (meters)	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Distance to nearest public transportation hub (meters)	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
Panel D: Components with controls			
Homicides	-0.021	0.021	0.000
	(0.059)	(0.067)	(0.044)
Robberies	0.001	0.002	-0.003
	(0.009)	(0.012)	(0.004)
Drug dealing	-0.003	0.056	-0.053
	(0.089)	(0.104)	(0.067)
Attempted homicides	0.130	-0.085	-0.045
	(0.094)	(0.112)	(0.058)
Illegal possession of a firearm	-0.086	0.020	0.066
	(0.070)	(0.122)	(0.099)
Number of buildings on block	0.001	-0.000	-0.001
	(0.001)	(0.001)	(0.001)
Area of block	-0.000	-0.000	0.000**
	(0.000)	(0.000)	(0.000)
Distance to nearest army battalion (meters)	0.000**	-0.000*	-0.000
	(0.000)	(0.000)	(0.000)
Distance to nearest police station (meters)	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Distance to nearest public transportation hub (meters)	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
Observations	1,254	1,254	1,254
Panel A: F-stat on index without controls	0.006	0.008	0.070
Panel B: F-stat on components without controls	1.166	0.383	0.504
Panel C: F-stat on index	0.008	0.093	0.516
Panel C: F-stat on index with controls	1.925*	1.232	1.529
Panel D: F-stat on components	0.785	0.217	0.509
Panel D: F-stat on components with controls	1.671*	0.804	1.089

Notes: Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.2: **Balance using endline survey data**

	Control	Treatment	Spillover
Panel A: Demographic controls			
Age	-0.00056	0.00092	-0.00035
	(0.00052)	(0.00066)	(0.00041)
Gender	0.00739	-0.00711	-0.00028
	(0.01256)	(0.01590)	(0.00980)
Education (years)	0.00118	-0.00122	0.00003
	(0.00200)	(0.00273)	(0.00170)
Panel B: Demographic and geo controls			
Age	-0.00010	0.00069	-0.00059
	(0.00047)	(0.00059)	(0.00038)
Gender	0.00602	-0.00648	0.00046
	(0.01244)	(0.01581)	(0.00983)
Education (years)	0.00329*	-0.00202	-0.00127
	(0.00181)	(0.00241)	(0.00149)
Number of buildings on block	-0.00007	0.00048	-0.00041
	(0.00082)	(0.00119)	(0.00085)
Area of block	-0.00000	-0.00001**	0.00001**
	(0.00000)	(0.00001)	(0.00000)
Distance to nearest army battalion (meters)	0.00004**	-0.00002	-0.00002
	(0.00002)	(0.00002)	(0.00002)
Distance to nearest police station (meters)	-0.00004	0.00001	0.00003
	(0.00003)	(0.00004)	(0.00002)
Distance to nearest public transportation hub (meters)	0.00007**	-0.00005	-0.00002
	(0.00004)	(0.00005)	(0.00003)
Observations	7,918	7,918	7,918
Panel A: F-stat on individual-level controls	1.484	1.708	0.462
Panel B: F-stat on block-level controls	2.060	1.680	0.824
Panel B: F-stat on individual- and block-level controls	1.609	1.321	1.867*

Notes: Standard errors, clustered by block, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.3: **Descriptive statistics for patrol activities**

	All blocks
Avg. # of patrols per block	5.06
	(5.13)
Avg. length of patrol (min.)	11.14
	(4.21)
Avg. # of soldiers per patrol	7.46
	(0.54)
Avg. % of patrols on correct block per night	80.26
	(0.17)
% of patrols with at least 1 stop and frisk	44.739
% of patrols with at least 1 ID check	7.685
% of patrols with at least 1 drug seizure	7.960
% of patrols with at least 1 arrest	0.183
% of patrols with at least 1 detention	0.091

Notes: Standard deviations in parentheses.

Figure A.1: **Alleged abuses by state security forces reported to Attorney-General's Office, September 30–November 18, 2019**

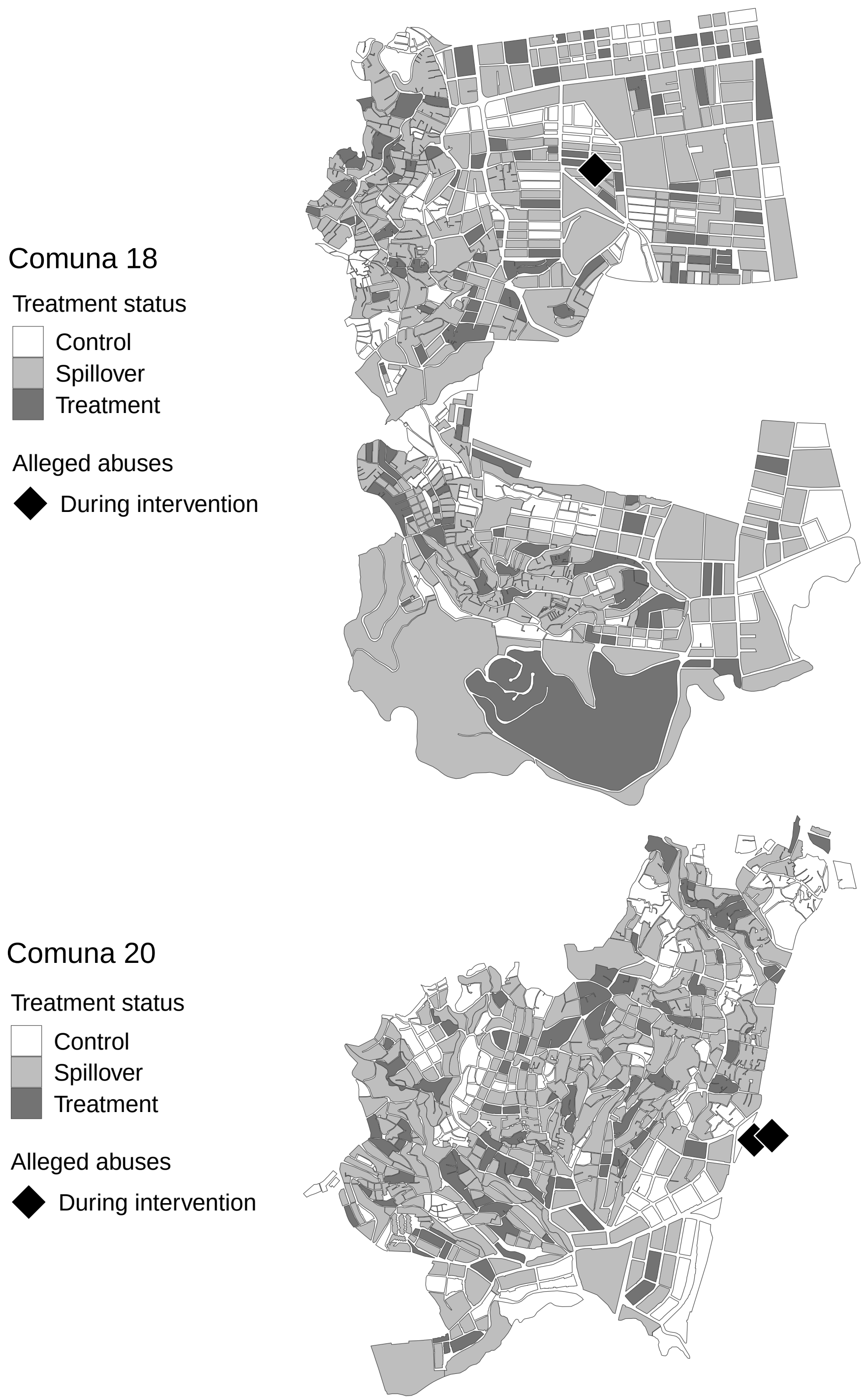


Table A.4: **Treatment effect on crime in administrative data using alternative approaches to estimating spillover effects**

	Crime index					
	Linear Decay		**Exponential Decay**		**Prop. Treated Blocks**	
	During intervention	After intervention	During intervention	After intervention	During intervention	After intervention
Treatment	0.023	0.061	0.023	0.059	0.034	-0.393
	(0.031)	(0.040)	(0.031)	(0.041)	(0.190)	(0.313)
Spillover	-0.007	-0.070*	-0.000	-0.096*		
	(0.018)	(0.033)	(0.040)	(0.048)		
Individual controls	✗	✗	✗	✗	✗	✗
Neighborhood FE	✓	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓	✓
Observations	1167	1167	1167	1167	202	202
R^2	0.33	0.48	0.33	0.48	0.46	0.72

Notes: ITT on additive indices of crime during (column 1) and after (column 2) the intervention based on administrative data, using alternative approaches to estimating spillover effects. Columns 1 and 2 use a linear decay function. Columns 3 and 4 use an exponential decay function. Columns 5 and 6 subsets to the treatment group and replaces the treatment and spillover indicators with the proportion of adjacent treatment blocks that are also assigned to treatment. All specifications include neighborhood fixed effects and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.5: **Treatment effects on crime, crime victimization, crime witnessing, and perceptions of safety with RI p-values and multiple comparisons correction**

	Admin data		Endline survey			
	Crime incidence		Crime victimization		Crime witnessing	Safety perceptions
	During intervention	After intervention	During intervention	After intervention	After intervention	After intervention
Treatment	0.003	0.110**	0.006	-0.007	0.153***	-0.050
	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
Spillover	-0.038	0.083*	0.026	0.013	0.186***	-0.066
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
Individual controls	✗	✗	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓	✓
Observations	1167	1167	7845	7845	7837	7707
R^2	0.33	0.48	0.03	0.03	0.12	0.09
Control mean	0.160	0.160	-0.021	-0.016	-0.119	0.077
RI p-value (treatment)	0.959	0.136	0.927	0.914	0.038	0.448
RI p-value (spillover)	0.411	0.138	0.610	0.802	0.001	0.199
qval-treatment	0.935	0.088	0.935	0.935	0.020	0.584
qval-spillover	0.319	0.176	0.467	0.729	0.001	0.201
BH-treatment		✓			✓	
BH-spillover					✓	
Holm-treatment					✓	
Holm-spillover					✓	

Notes: ITT on additive indices of crime during (column 1) and after (column 2) the intervention based on administrative data, and standardized additive indices of crime victimization during (column 3) and after (column 4) the intervention, crime witnessing after the intervention (column 5), and perceptions of safety after the intervention (column 6) in the endline survey data. All specifications include neighborhood fixed effects and block-level controls. Columns 1 and 2 include a lagged dependent variables; columns 3-6 include individual-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are in parentheses, and are clustered by block in columns 3-6. *** p<0.01, ** p<0.05, * p<0.1. We report randomization inference p-values for each ITT estimate. We also report the Benjamini-Hochberg q-value for each estimate, and indicate whether each p-value falls below the corresponding q-value. We also indicate whether each p-value falls below the Holm-Bonferroni threshold. We assume that each model amounts to one hypothesis test.

Table A.6: **Treatment effects on human rights abuses with RI p-values and multiple comparisons correction**

	Monitoring data		Survey data	
	Police abuse	**Military abuse**	**Police abuse**	**Military abuse**
Treatment	0.037***	0.010*	0.011	-0.001
	(0.011)	(0.006)	(0.013)	(0.004)
Spillover	0.016	0.001	0.030**	0.002
	(0.011)	(0.003)	(0.012)	(0.003)
Individual-level controls	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓
Observations	1970	1970	7908	7908
R^2	0.05	0.05	0.05	0.01
Control mean	0.015	0.000	0.114	0.012
RI p-value (treatment)	0.103	0.430	0.541	0.783
RI p-value (spillover)	0.385	0.942	0.028	0.613
qval-treatment	0.003	0.142	0.556	0.716
qval-spillover	0.275	0.778	0.044	0.726
BH-treatment	✓			
BH-spillover			✓	
Holm-treatment	✓			
Holm-spillover			✓	

Notes: ITT on abuses by police (columns 1 and 3) and military (columns 2 and 4) based on monitoring (columns 1 and 2) and survey data (columns 3 and 4). All specifications include neighborhood fixed effects and individual- and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors, clustered by block, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We report randomization inference p-values for each ITT estimate. We also report the Benjamini-Hochberg q-value for each estimate, and indicate whether each p-value falls below the corresponding q-value. We also indicate whether each p-value falls below the Holm-Bonferroni threshold. We assume that each model amounts to one hypothesis test.

Table A.7: **Treatment effects on perceptions of the police and military and demand for military policing with RI p-values and multiple comparisons correction**

	Survey data		**Behavioral data**
	Perceptions of police	**Perceptions of military**	**Demand for military**
Treatment	0.030	0.079*	0.053**
	(0.043)	(0.044)	(0.023)
Spillover	-0.021	-0.002	0.033**
	(0.034)	(0.036)	(0.016)
Individual-level controls	✓	✓	✗
Block-level controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Observations	7794	7757	1167
R^2	0.05	0.04	0.11
Control mean	-0.001	-0.014	0.028
RI p-value (treatment)	0.610	0.193	0.175
RI p-value (spillover)	0.644	0.966	0.290
qval-treatment	0.478	0.105	0.068
qval-spillover	0.796	0.958	0.136
BH-treatment			✓
BH-spillover			
Holm-treatment			✓
Holm-spillover			

Notes: ITT on standardized additive indices of perceptions of police (column 1) and military (column 2) in the survey and demand for military policing (column 3) in our behavioral measure. Columns 1 and 2 include neighborhood fixed effects and individual- and block-level controls. Column 3 includes neighborhood fixed effects and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors, clustered by block in columns 1 and 2, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We report randomization inference p-values for each ITT estimate. We also report the Benjamini-Hochberg q-value for each estimate, and indicate whether each p-value falls below the corresponding q-value. We also indicate whether each p-value falls below the Holm-Bonferroni threshold. We assume that each model amounts to one hypothesis test.

Table A.8: **Treatment effects on support for vigilantism and military coups with RI p-values and multiple comparisons correction**

	Survey data			
	Support for military coups		Support for vigilantism	
	In response to crime	In response to corruption	Bypassing the legal system	Bypassing the police
Treatment	0.048**	0.014	0.086**	-0.014
	(0.020)	(0.020)	(0.040)	(0.043)
Spillover	0.048***	0.007	0.061*	0.006
	(0.017)	(0.017)	(0.035)	(0.035)
Individual-level controls	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓
Observations	7806	7830	7882	7868
R^2	0.04	0.04	0.03	0.03
Control mean	0.474	0.664	-0.032	0.020
RI p-value (treatment)	0.074	0.599	0.101	0.813
RI p-value (spillover)	0.021	0.767	0.138	0.900
qval-treatment	0.068	0.623	0.068	0.753
qval-spillover	0.025	0.869	0.167	0.869
BH-treatment	✓		✓	
BH-spillover	✓			
Holm-treatment	✓			
Holm-spillover	✓			

Notes: ITT on dummies indicating support for military coups in response to crime (column 1) and corruption (column 2), and on indices of support for bypassing the legal system (column 3) and the police (column 4) to punish criminals based on survey data. All specifications include neighborhood fixed effects and individual- and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors, clustered by block, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We report randomization inference p-values for each ITT estimate. We also report the Benjamini-Hochberg q-value for each estimate, and indicate whether each p-value falls below the corresponding q-value. We also indicate whether each p-value falls below the Holm-Bonferroni threshold. We assume that each model amounts to one hypothesis test.

Table A.9: **Treatment effects on turnout for right-wing candidates and support for and participation in left-wing protests with RI p-values and multiple comparisons correction**

	Survey data		
	Turnout for conservative candidates	**Attitudes and behavior towards protests**	
	All	Approves	Participated
Treatment	-0.012	-0.077*	-0.005
	(0.010)	(0.042)	(0.011)
Spillover	0.004	-0.008	-0.007
	(0.009)	(0.036)	(0.010)
Individual-level controls	✓	✓	✓
Block-level controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Observations	7908	7419	7498
R^2	0.02	0.09	0.07
Control mean	0.077	-0.001	0.094
RI p-value (treatment)	0.380	0.150	0.765
RI p-value (spillover)	0.702	0.846	0.568
qval-treatment	0.355	0.210	0.691
qval-spillover	0.825	0.825	0.825
BH-treatment			
BH-spillover			
Holm-treatment			
Holm-spillover			

Notes: ITT on turnout for right-wing candidates (column 1) and support for (column 3) and participation in (column 4) left-wing protests based on survey data. All specifications include neighborhood fixed effects and individual- and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors, clustered by block, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We report randomization inference p-values for each ITT estimate. We also report the Benjamini-Hochberg q-value for each estimate, and indicate whether each p-value falls below the corresponding q-value. We also indicate whether each p-value falls below the Holm-Bonferroni threshold. We assume that each model amounts to one hypothesis test.

Table A.10: **Treatment effect on human rights abuses in monitoring data disaggregated by date of survey**

	Monitoring data							
	Before November 19		**After November 19**		**Before December 2**		**After December 2**	
	Police abuse	**Military abuse**	**Police abuse**	**Military abuse**	**Police abuse**	**Military abuse**	**Police abuse**	**Military abuse**
Treatment	0.034*	0.000	0.043***	0.020*	0.029*	0.001	0.037**	0.018
	(0.020)	(.)	(0.014)	(0.011)	(0.017)	(0.001)	(0.016)	(0.014)
Spillover	0.002	0.000	0.031**	0.002	0.002	-0.003	0.028**	0.001
	(0.022)	(.)	(0.012)	(0.005)	(0.018)	(0.003)	(0.013)	(0.006)
Individual-level controls	✓	✓	✓	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	829	829	1141	1141	1075	1075	895	895
R^2	0.08	.	0.05	0.05	0.07	0.04	0.06	0.05
Control mean	0.022	0.000	0.011	0.000	0.024	0.000	0.009	0.000

Notes: ITT on abuses committed by police and military based on monitoring data during (columns 1-4) and after (columns 5-8) the intervention. In columns 1, 2, 5, and 6 we assume that any abuses reported by respondents who were surveyed before November 19 occurred during the intervention, and that any abuses reported by respondents who were surveyed after November 19 occurred after the intervention. In columns 3, 4, 7, and 8 we assume that any abuses reported by respondents who were surveyed before December 2 occurred during the intervention, and that any abuses reported by respondents who were surveyed after December 2 occurred after the intervention. All specifications include neighborhood fixed effects and individual- and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors, clustered by block, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.11: **Treatment effect on police and military presence and arrests**

	Monitoring data				**Survey data**			
	During intervention				**After intervention**			
	Seen on block		**Seen making arrests**		**Seen on block**		**Seen making arrests**	
	Police	Military	Police	Military	Police	Military	Police	Military
Treatment	0.032	0.105***	0.033**	0.011*	0.044**	0.023	0.059***	0.011
	(0.041)	(0.036)	(0.014)	(0.006)	(0.019)	(0.020)	(0.022)	(0.011)
Spillover	0.015	0.079*	0.036**	0.004	0.034**	0.005	0.062***	0.019**
	(0.047)	(0.043)	(0.017)	(0.004)	(0.016)	(0.018)	(0.018)	(0.009)
Individual-level controls	✓	✓	✓	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1970	1970	1970	1970	7908	7908	7908	7908
R^2	0.07	0.10	0.05	0.04	0.04	0.13	0.05	0.02
Control Mean	0.612	0.257	0.034	0.000	0.788	0.387	0.302	0.065

Notes: ITT on police and military presence and arrests during the intervention based on monitoring survey (columns 1-4) and after the intervention based on endline survey (columns 5-8). All specifications include neighborhood fixed effects and individual- and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors, clustered by block, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.12: **Treatment effect on cooperation with authorities**

	Survey data		
	All	Police	Military
Treatment	0.086**	0.080*	0.069*
	(0.042)	(0.043)	(0.040)
Spillover	0.087**	0.091**	0.034
	(0.034)	(0.036)	(0.029)
Individual-level controls	✓	✓	✓
Block-level controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Observations	7824	7847	7866
R^2	0.04	0.04	0.02
Control mean	-0.052	-0.051	-0.030

Notes: ITT on cooperation with the authorities in general (column 1) and the police (column 2) and military (column 3) separately based on endline survey. All specifications include neighborhood fixed effects and individual- and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors, clustered by block, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.