Military Policing Exacerbates Crime and Human Rights Abuses: A Randomized Controlled Trial in Cali, Colombia*

Robert A. Blair†
Michael Weintraub‡

Abstract

Governments across the developing world rely on their armed forces for domestic policing operations. Advocates of these "*mano dura*" (iron fist) policies view them as necessary to control violent crime, while detractors claim they undermine human rights. We experimentally evaluate a military policing intervention in Cali, Colombia, the country's third largest city and among its most violent. The intervention involved recurring, intensive military patrols targeting crime hot spots, randomly assigned at the city block level. Using administrative crime and human rights data, surveys of more than 10,000 Cali residents, and detailed firsthand observations from civilian monitors, we find that military policing had weak (if any) effects on crime while the intervention was ongoing, and adverse effects after it was complete. We observe higher rates of crime, crime witnessing, and crime reporting in the weeks after the intervention, combined with higher rates of arrests. We also find some suggestive evidence of increased human rights abuses, though these appear to have been committed primarily by police officers rather than soldiers. Our results suggest that the benefits of military policing are small and not worth the costs, and that governments should seek other ways to control crime in the world's most violent cities.

**Acknowledgements:* Funding for this project was provided by the UK Foreign, Commonwealth & Development Office and 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 (Andrés Villamizar, Pablo Uribe, Alberto Sánchez, Pablo Gracia, María Alejandra Arboleda and Juan Diego Tabares) for logistical oversight; the Third Battalion of the Armed Forces of Colombia (especially Colonel Omar Arciniegas) for its commitment to the evaluation; and Sebastián Hernández, Carlos Gúzman, Carlos Bohm, Abraham Farfán, Sergio Flórez, and María Fernanda Cortés for terrific research assistance. For helpful feedback we thank Peter Aronow, Chris Blattman, Alex Coppock, Erica De Bruin, Leopoldo Fergusson, Don Green, Andrés Moya, Sarah Parkinson, Santiago Tobón, Juan Fernando Vargas, Jess Zarkin, and participants at workshops and conferences convened by the Folke Bernadotte Academy, Uppsala University, the *Centro de Estudios sobre Seguridad y Drogas* (CESED) at Universidad de los Andes, and the American Political Science Association. The Ethics Committee at Universidad de los Andes approved this project (Acta 1073 de 2019 and Acta 1034 de 2019).
†Assistant 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

A strict separation between the military and the police has long been viewed as essential for transitions to democracy and democratic consolidation [1–3]. Across the Global South, however, democratically elected governments routinely rely on their armed forces for domestic policing operations. This "*mano dura*" (iron fist) approach to law enforcement is especially pervasive in Latin America, the world's most violent region [4]. The governments of Bolivia, Brazil, Colombia, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Paraguay, and Peru have all "constabularized" their militaries to varying degrees [5]. Military policing is common outside of Latin America as well, for example in Indonesia [6], the Philippines [7], and South Africa [8]. Even in the US, with its long history of opposition to military involvement in domestic policing operations, commentators have on occasion urged the deployment of troops to support police departments in "high-crime, drug-infested urban areas" [9, p. 220].

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 experienced a homicide rate of 46.7 per 100,000 residents, nearly double the rate of Colombia's second largest city (Medellín) and more than triple the rate of the capital (Bogotá). 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.[1] Military policing programs like *Plan Fortaleza* are a defining feature of *mano dura* policies in Latin America and beyond [10]. 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.

Advocates view military policing as a necessary temporary measure to reduce crime and improve citizen perceptions of security [11]. Opponents view it as a threat to human rights [12]. Yet despite the increasing prevalence of military policing throughout the Global South, empirical evidence of its efficacy remains scarce. There is a large and established literature on hot spots policing, much of it conducted in the Global North, but these studies involve policing by actual police forces rather than militaries, and typically do not address militarization (see [13, 14] for

[1] *Comunas* are the highest administrative unit in Cali. *Comunas* are divided into *barrios* (neighborhoods), which are further divided into *manzanas* (blocks).

reviews). Most studies of militarization focus on SWAT team deployments or transfers of military hardware to police departments in the US, with mixed results [15–20]. Other studies explore militarization of the police in Latin America, but do not address policing by the military per se [21–23].

This is a subtle but crucial distinction. Militaries tend to differ dramatically from even the most heavily militarized police forces in the type and intensity of training they receive; in their more inflexible vertical command structures; in the legal frameworks within which they operate; and, most important, in the expectation that they will use deadly force not to "serve and protect" civilians, but rather to "overwhelm and defeat" enemies on the battlefield [24, p. 329]. Both supporters and detractors of military policing cite these differences as justification for their respective positions. As one prominent critic warns, "no one should suffer the illusion that military forces could ever execute the laws with the same sensitivity to civil liberties as regular police forces," because to do so is "at odds with the central imperatives of military service" [9, p. 227].

To date, however, claims such as these remain almost entirely anecdotal and impressionistic. Only a small handful of studies have tested the effects of constabularizing the military for purposes of law enforcement, all using observational data [5, 25, 26]. These studies have been informative and pathbreaking. But observational research on military policing must overcome enormous inferential challenges. Soldiers are typically only deployed to conduct domestic policing operations where crime and violence have become unmanageable for police departments [11]. This makes it difficult to disentangle the effects of military policing from the effects of the conditions that justified (or were believed to justify) military policing in the first place. Quasi-experimental evidence on the efficacy of military policing is rare, and experimental evidence is, to our knowledge, non-existent.

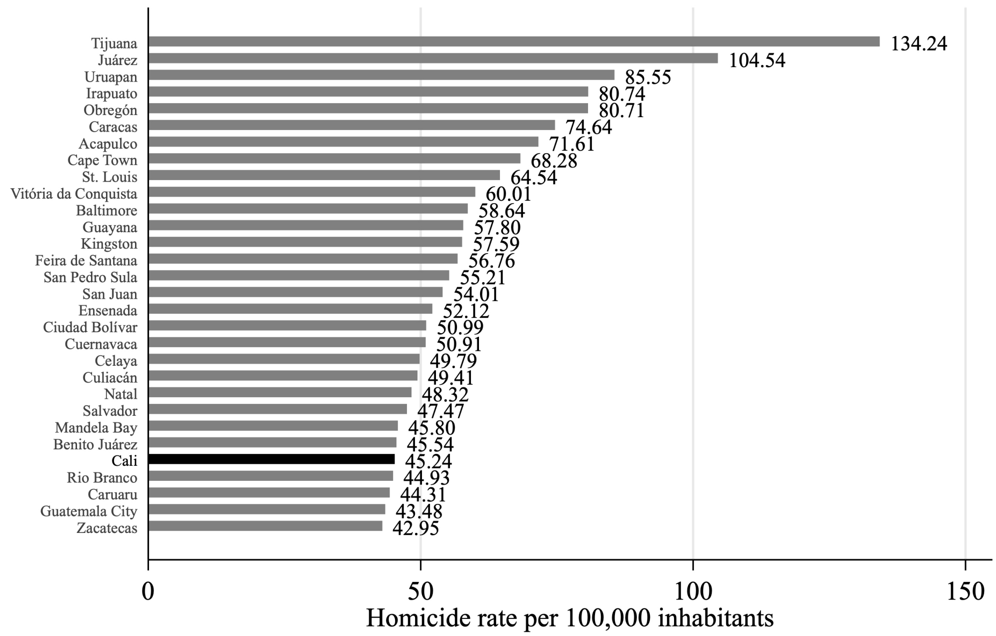
We evaluate Cali's *Plan Fortaleza* program using a randomized controlled trial, with treatment assigned at the level of the *manzana* (city block). Our sample consists of 1,255 blocks in total, 214 of which were assigned to treatment. Another 765 blocks that were adjacent to at least one treatment block were assigned to a spillover group; the remaining 275 blocks were assigned to control. (We model more complex spillover dynamics in Appendix E.) To test the effects of the program, we combine timestamped, geolocated administrative data on crime and human rights abuses with two waves of original surveys reaching over 10,000 respondents in total. Our first wave of survey data was collected continuously throughout the intervention; the second was collected roughly three months after the intervention was complete, allowing us to test effects both during and after implementation. We complement these data with detailed firsthand observations from civilian monitors hired to accompany the soldiers while on patrol.

We find that *Plan Fortaleza* had little (if any) impact on crime in the administrative data while the intervention was ongoing. The effects are weak and sensitive to specification, and are confined to days and times when soldiers were physically present on the streets. In contrast, we observe a substantively large and statistically significant *increase* in crime in the weeks after the intervention was complete, along with corresponding increases in citizens' accounts of (1) witnessing and (2) reporting crimes to the authorities, (3) the intensity of police presence, and (4) the rate of arrests on treatment and spillover blocks. The adverse effects on crime do not appear to be artifacts of changes in the distribution of police officers across blocks, or of heightened vigilance on the part of civilians. Perhaps relatedly, we also find no evidence of improved perceptions of safety except (possibly) among business owners.

Finally, we find some evidence of increased human rights abuses, but only in the survey data. In most (but not all) cases these abuses appear to have been perpetrated by police officers rather than soldiers, possibly in the course of effecting arrests. (Colombian soldiers can detain but cannot arrest suspected criminals.) This suggests a potential adverse unintended consequence of military policing that critics rarely mention: if soldiers and police officers "co-produce" security, then military involvement in law enforcement may exacerbate abuses even if the military itself is not responsible for committing them. But this interpretation is suggestive, as we find no evidence of increased abuses in administrative data collected from the Office of the Attorney-General, nor in the firsthand observations of civilian monitors.

Taken together, our results suggest that governments should seek other strategies for curbing

Figure 1: **The world's most dangerous cities by homicide rate, 2019**



Notes: Data on homicides is from the NGO Seguridad, Justicia, y Paz.

crime in the world's most violent cities. As with any study focused on a single case, we cannot be sure whether these results will generalize to other contexts. Cali does, however, share traits with other urban settings in Latin America, and the intervention we study resembles existing military policing programs elsewhere. Like other cities in Colombia (and cities in other Latin American countries), Cali features extreme inequality across space: access to security and other services depends on where one lives and the socioeconomic class to which one belongs. The police in Cali, as in other Latin American cities, have a checkered past that has featured corruption, collusion with armed groups, and human rights abuses, which helps explain widespread popular support for military policing. The program we evaluate is place-based, analogous to hot spots policing interventions commonly undertaken by police departments in the US and elsewhere, including Colombia [27, 28]. Almost all existing studies of hot spots policing focus on particular neighborhoods within particular cities [14]. Our study of military policing follows a similar design.

Cali is also an important test case for the efficacy of military policing. Many residents arrived in Cali after fleeing violence in other areas of Colombia's Pacific Coast, and while crime has declined since the worst days of violence between the Cali and Medellín cartels in the 1990s—

when the city's homicide rate reached nearly 125 per 100,000 residents—Cali remains among the most dangerous cities in the world, as Figure 1 illustrates. Gangs and other armed actors compete for local extortion and money laundering opportunities, as well as access to drug trafficking routes through the nearby port of Buenaventura. The quality of administrative crime data is also unusually high in Cali, a result of close coordination among the National Police, the Office of the Attorney-General, the Coroner's Office, and other civilian authorities. Given ongoing struggles to reduce crime, Cali is precisely the sort of setting where military policing might be—and has been—expected to help. Lessons learned from Cali can and should inform debates in cities elsewhere in Latin America (and potentially beyond) where military policing is routinely used but seldom rigorously evaluated.

THE LOGIC OF MILITARY POLICING

Military policing is motivated by the belief that soldiers can deter and incapacitate criminals more effectively than police officers. Soldiers typically have better communication, transportation, and logistical capacity than police officers, and must undergo more (and more rigorous) training [3, 26]. They also have greater coercive capacity. This may be especially advantageous in Latin America, where much crime can be traced to gangs, cartels, and other organized criminal syndicates, which have proven to be "daunting foes" for the police [29, p. 104]. Many Latin American police forces are "poorly trained, poorly funded, and even complicit with organized crime" [5, p. 5]. Advocates cite these differences to argue that soldiers should be more effective at both deterrence (i.e. preventing future criminals from committing crimes) and incapacitation (i.e. removing perpetrators of past crimes from the streets), and that military policing should thus reduce the objective prevalence of crime [30, 31].

Advocates also believe that military policing should improve subjective perceptions of safety among citizens and business owners. In many developing countries, and in Latin America in particular, citizens perceive the military to be more effective and better trained than the police

[11]. They also tend to express greater trust in the military. For example, of the 25 North, South, Central American, and Caribbean countries surveyed in the 2014 Americas Barometer survey, respondents expressed greater trust in the military than the police in all but one country (Chile), typically by double digit margins. Other than the Catholic Church, no other institution is as widely trusted as the military in Latin America [29]. If citizens and business owners view the military as more effective, better trained, and more trustworthy than the police, then they should feel safer in the presence of military policing.

Finally, advocates argue that because soldiers are better trained and subject to more stringent accountability mechanisms, they should be more respectful of human rights than police officers [31]. Again, this distinction may be especially salient in Latin America, where indiscipline among the region's police forces has manifested in widespread illegality and extrajudicial violence [32]. Latin American citizens tend to perceive the military as more respectful of human rights than the police [11], even in settings where soldiers are engaged in domestic law enforcement (and where their actions are therefore more visible to civilians). Soldiers also typically operate under stricter hierarchies of command and control [26], which may reduce the risk that misconduct goes undetected and unpunished. This line of reasoning has led some to conclude that "maybe the police aren't militarized enough" [33].

In our pre-analysis plan (PAP)[2] we hypothesized that military policing would reduce the objective prevalence of crime and improve subjective perceptions of safety without exacerbating human rights abuses. But all three of the arguments described above are hotly contested. Opponents of military policing counter that deterrence and incapacitation depend on swift and effective criminal investigation [34], which soldiers are not trained to do. Gangs and other organized criminal syndicates may respond strategically to the military's threat of deadly force by engaging in yet more violence themselves [35], thus exacerbating crime and diminishing citizen perceptions of security [25, 26]. Perhaps most important, opponents warn that military policing replaces the

[2]Our PAP was filed with the Evidence in Governance and Politics (EGAP) network prior to endline data collection. Anonymized copies of our PAP and associated updates are included as appendices. We document all deviations from the PAP in Appendix F.

"traditional police orientation of 'protect and serve"' with a "military orientation of 'overwhelm and defeat"' [24, pp. 329–30]. Soldiers are "conditioned by years of rigorous training and indoctrination," and may be "hard-wired" to use excessive force in ways that police officers are not [11, p. 9], with potentially adverse consequences for human rights.

THE *Plan Fortaleza* PROGRAM

The evidence underlying these arguments is almost entirely anecdotal. Rigorously testing the efficacy of military policing is of urgent theoretical and practical concern, especially given the continued proliferation of *mano dura* policies across the Global South. We contribute to this debate with an experimental evaluation of the *Plan Fortaleza* program in Cali, Colombia. The program consisted of recurring, intensive vehicular and foot patrols by heavily armed soldiers from the Military Forces of Colombia (FMC). Cali has 22 *comunas* in total; *Plan Fortaleza* focused on *comunas* 18 and 20, both hot spots for crime, as we show in Figure A.5 in the appendix. The two *comunas* comprise 30 *barrios*, or neighborhoods. Their combined population was approximately 215,000 at the time of our study.

Patrols were implemented by soldiers from two units of the FMC: the Military Police and the Special Forces. While both units consist entirely of soldiers, Special Forces tend to be older, have more field experience (including in combat with guerrilla groups), and use more advanced military hardware. To minimize logistical problems, the two units never patrolled the same *comuna* on the same day; instead, they alternated following a 12-day rotation schedule, illustrated in Table A.1 in the appendix. Each patrol consisted of six to eight soldiers from one of these two units, with seven to eight teams patrolling more or less simultaneously every weekday night. While on patrol, soldiers checked IDs and business licenses, searched residents for possession of drugs and weapons ("*requisas*"), erected road blocks, detained suspected criminals, and conversed with residents. All patrols occurred between the hours of 5:00pm and midnight, Monday to Friday. All blocks also had some police presence.

Our unit of randomization is the *manzana* (city block). Each treatment block was assigned to receive 30 minutes of military patrols roughly every six days. In reality, the average time spent patrolling was around 11 minutes per block per day, due in part to the small size of most blocks and the large number of soldiers on patrol.[3] Since all patrols originated from the same battalion, and since we did not specify the routes the soldiers should take to reach each treatment block, we recognized at the outset that the probability of spillover would be high. We discuss this in more detail below, and model spillover in our analyses. 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 around the city and country.

While the intervention was relatively short, it was not atypical of the way Latin American militaries often participate in law enforcement. Even in countries undergoing "generalized constabularization of the military," soldiers usually participate in temporally and geographically delimited operations targeting particular areas characterized by high rates of violent crime and/or drug trafficking [36, p. 526]. Permanent or semi-permanent military occupations are less common, though they do occur, as in Mexico following President Felipe Calderón's declaration of war against drug cartels in 2006 [5]. We cannot be certain whether our findings might generalize to these latter situations, nor can we be sure how they might have differed if *Plan Fortaleza* had been longer (or shorter) in duration.

Comunas 18 and 20 are densely populated and difficult to navigate. In some parts of *comuna* 20, for example, streets are unlit alleys that connect to roads via steep, concrete stairs. To help guide the soldiers, local civilian monitors accompanied each patrol. Monitors used GPS devices and smartphones equipped with a customized Google Maps interface to direct soldiers to their assigned treatment blocks. We provide examples of this interface in Figures A.1 and A.2 in the appendix. The monitors also used smartphones to collect data on soldiers' operations during the patrols. To track treatment compliance, we established geo-fences of 25 meters around each treatment block and calculated the time that each patrol spent within its assigned geo-fence. We provide descriptive

[3] The average perimeter for blocks in our sample is 283 meters, with a standard deviation of 248 meters.

statistics on the patrols in Table A.2 in the appendix, and discuss our safety protocols and the possibility of Hawthorne effects induced by the monitors' presence in further detail below.

Research Design

Randomization

The 30 neighborhoods in our sample consist of 1,255 city blocks, with an average of 42 blocks per neighborhood. We stratified by neighborhood, then randomized such that approximately 1/6 of all blocks in each neighborhood were assigned to treatment.[4] We 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 B. We assigned all remaining blocks to the control group. Our sample thus consists of 214 treatment blocks, 765 spillover blocks, and 275 control blocks. We provide power calculations in Appendix C and balance tests in Appendix D. In Appendix E we explore different ways of modeling spillover effects, including linear and exponential decay.

Data

We collected data from four sources. First, we collected timestamped, geocoded administrative data on crime, including homicides, robberies, illegal drug sales, and illegal possession of firearms. These data span a period beginning nine months before the intervention (January 1, 2019) and ending six weeks after (December 31, 2019). The quality of administrative crime data is unusually high in Cali, where representatives of the Mayor's Office meet weekly 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 timestamped, geocoded administrative data on human rights abuses from the Office of the Attorney-General, which is responsible for

[4] We gain no additional statistical power for estimating treatment effects and only marginal statistical power for estimating spillover effects when we increase the proportion of treated blocks above 1/6.

investigating and prosecuting police and military misconduct. The Attorney-General's data consist of alleged abuses reported by victims and witnesses, and again cover all of 2019.

Second, we conducted an original household survey of 2,096 randomly selected residents of the two *comunas* in our sample between October 17 and December 19, 2019, beginning while the intervention was ongoing and continuing for roughly a month after it ended. We surveyed three residents and two business owners on each of 416 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 our monitoring survey. Third, we conducted another original household survey of 7,921 randomly selected residents and business owners between January 17 and February 25, 2020, between two and three months after the end of the intervention. On average we surveyed six residents and two business owners per block.[5] We refer to this as our endline survey.[6]

Finally, we collected GPS data and detailed firsthand observations from the civilian monitors hired to accompany the soldiers while on patrol. Because we only have these data for the treatment group, we do not use them to estimate treatment effects; instead, we use them to measure the duration of each patrol, the number of soldiers on each patrol, and the soldiers' activities while on patrol, including any acts of verbal or physical abuse. To minimize Hawthorne effects, monitors were instructed to be as discreet as possible when documenting soldiers' activities, and to remain in the patrol vehicle at all times. To facilitate discretion and standardize data collection, monitors used a smartphone app developed for this project to record their observations. The same monitors accompanied the same patrols repeatedly for nearly two months, allowing the soldiers to acclimate to their presence, thus further mitigating the risk of Hawthorne effects. 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—

[5]Because of differing sampling strategies across the two surveys, measurements were taken from distinct samples.

[6]In addition to crime victimization, crime witnessing, perceptions of safety, and experiences of abuse, the endline survey included a variety of questions on perceptions of the police and military, political beliefs, and voting behavior. For compactness we focus in this paper on crime, perceptions of safety, and abuses, as these are the outcomes of most urgent concern to both proponents and detractors of military policing. We report treatment effects on perceptions, political beliefs, and voting behavior in a separate paper.

even without the monitors. Nonetheless, our results should be interpreted as capturing the effects of military patrols in the presence of civilian monitors.

ESTIMATION

Our unit of randomization is the city block. Some outcomes are measured at the individual level using survey data, others at the block level using administrative data. Following our PAP, we estimate the intention-to-treat (ITT) effect of the *Plan Fortaleza* program using a weighted least squares 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, we cannot calculate inverse probability weights (IPWs) analytically. Instead, we bootstrap our randomization procedure and estimate the probability that each block is assigned to the treatment, spillover, and control group across 1,500 replications. We use these estimates to generate IPWs. Because of the way blocks are distributed across neighborhoods, some (though very few) have a 0 probability of assignment to the spillover or control group.[7] We exclude blocks with 0 probability of assignment to control when estimating the ITT [37]. 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.

When testing treatment and spillover effects at the block level, we estimate

$$y_{jk} = \theta t_{jk} + \lambda s_{jk} + \beta \mathbf{X_{jk}} + \alpha_k + \epsilon_{jk}$$

where y_{jk} denotes the outcome for block j in neighborhood k; t_{jk} denotes assignment to treatment; s_{jk} denotes assignment to spillover; $\mathbf{X_{jk}}$ denotes block-level covariates;[8] α_k denotes neighborhood fixed effects; and ϵ_j is a block-level error term. When testing treatment and spillover effects at the

[7] Across the 30 neighborhoods, there is one block with 0 probability of assignment to spillover and another three blocks with 0 probability of assignment to control.

[8] Following our PAP, we control for area, distance to the nearest police station, distance to the nearest military battalion, and distance to the nearest public transportation hub based on administrative data. We also control for the average age and average years of education of residents on each block, and the percentage of men on each block, aggregating our individual-level survey data up to the block level.

individual level, we instead estimate

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

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; $\mathbf{Z_{ijk}}$ denotes individual-level covariates;[9] α_k denotes neighborhood fixed effects; and ϵ_{ijk} is an individual-level error term, clustered by block. There were five blocks (0.3% of the sample) where it was impossible to administer our endline survey due to safety concerns. We drop these blocks from our analysis.[10] We report results with multiple comparisons corrections in Appendix G, and heterogeneous treatment effects by gender and baseline crime rate in Appendix J.

SPILLOVER

Our research design allows us to estimate both direct and spillover effects of the *Plan Fortaleza* program. Criminologists distinguish between two types of spillover: displacement (whereby increased police presence displaces crime from one location to another nearby) and diffusion of benefits (whereby increased police presence in one location reduces crime in nearby locations). The literature on these possibilities is extensive; while results are mixed, the most recent research suggests that displacement tends to be minimal, that it is usually offset by treatment effects, and that diffusion of benefits is more common [38]. In Appendix E we test for the possibility of more complex spillover dynamics, including linear decay, exponential decay, and saturation.[11] We find little evidence of these dynamics at work in our sample.

[9] Again following our PAP, we control for age, gender, years of education, and years living on the block.

[10] In our PAP we proposed computing Lee bounds around our treatment effect estimates in order to account for attrition. Given that attrition was so minor, we omit this analysis here.

[11] In our PAP we also prespecified that we would model more complex spillover dynamics by estimating the "marginalized individualistic response," following **?**. We decided not to do this because the procedure is new and untested.

Compliance

Data collected by the civilian monitors suggest that treatment compliance was reasonably high, especially given the difficulty of navigating these neighborhoods. On any given night, soldiers correctly patrolled between 85% and 100% of treatment blocks on the randomization schedule. As a manipulation check, in Table A.30 in the appendix we show that residents of treatment and (to a lesser extent) spillover blocks were much more likely to report military presence (but no more or less likely to report police presence) than residents of control blocks during the intervention. This is as expected.

Ethics

Given the increasing prevalence of military policing in Colombia and across the Global South, the absence of evidence on its efficacy, and the arguments of advocates (including in the Colombian government) that military policing is necessary to curb violent crime, we believed a rigorous impact evaluation was essential to inform policymaking. The *Plan Fortaleza* program predated our study, and would have continued with or without our evaluation of it. Colombian municipal authorities and military officials had already selected the *comunas* and neighborhoods in which the intervention would occur; we randomized only the specific blocks where soldiers would and would not patrol.

Nonetheless, both the program and our evaluation of it posed several potential risks, which we sought to anticipate and mitigate. First, there was a risk that military patrols would subject residents to human rights abuses by soldiers. We sought to minimize this risk by using the firsthand observations of civilian monitors to document any abuses they observed. We maintained a direct line of communication with the Security and Justice Secretariat of Cali, which oversees military operations in the city, in order to report abuses in real time. As discussed below, the monitors recorded only one minor incident of verbal abuse and no incidents of physical abuse throughout the duration of the study.

Second, there was a risk that military patrols would subject civilians to violence by shifting the equilibrium distribution of gang presence and activity in the two *comunas* in our sample. 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 in Cali, given the city's highly fragmented landscape of organized crime. We saw no reason to expect the intervention to create a new, more violent equilibrium where none existed before. Prior to our study, the military (non-randomly) varied its patrol routes from day to day to prevent criminals from adapting to its presence; 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 civilians would face reprisals for participating in our monitoring or endline surveys, or that enumerators would face reprisals for administering the surveys. To minimize this risk, 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. Enumerators received specialized training and followed strict security protocols on all blocks, including a requirement to complete data collection by noon each day. 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 identify the civilian monitors, potentially subjecting them to harassment or violence. To minimize this risk, recruitment focused on monitors who did not live in the two *comunas* in our study, reducing the probability that they would be identified. Monitors also had a direct line of communication to the military and the Security and Justice Secretariat, which they could use to seek help if they suspected they were being watched or followed. To increase discretion and mitigate other potential risks to their safety, monitors were instructed to remain in the back of the patrol vehicle at all times. As additional precautions, moni-

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

	Admin data		Survey data		
	Crime incidence		**Crime victimization**		**Crime witnessing**
	During intervention	After intervention	During intervention	After intervention	After intervention
Treatment	0.003	0.110**	0.006	-0.007	0.153***
	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)
Spillover	-0.038	0.083*	0.026	0.013	0.186***
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)
Individual controls	✗	✗	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓
Observations	1167	1167	7845	7845	7837
R^2	0.33	0.48	0.03	0.03	0.12
Control mean	0.160	0.160	-0.021	-0.016	-0.119

Notes: ITT on crime during (column 1) and after (column 2) the intervention based on administrative data; crime victimization during (column 3) and after (column 4) the intervention based on survey data; and crime witnessing after the intervention (column 5) based on survey data. All specifications include neighborhood fixed effects and block-level controls. Models 3-6 also include individual-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.

tors were also provided with bulletproof vests and armbands clearly identifying them as civilians, thus reducing 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.

RESULTS

CRIME, CRIME VICTIMIZATION, AND CRIME WITNESSING

Table 1 reports the ITT of the *Plan Fortaleza* program on additive indices of crimes committed during (column 1) and after (column 2) the intervention using the administrative data. The index comprises murders, robberies, illegal drug sales, and illegal possession of weapons.[12] We find

[12] In our PAP we proposed to weight this index by the average prison sentence associated with each crime under Colombian law, following an approach that appeared in an early draft of [27]. This approach is not standard in the literature, yields results that are difficult to interpret substantively, and was abandoned in later drafts of [27]. We focus on results using an unweighted additive index here, and report results using the weighted index in Appendix

no evidence that *Plan Fortaleza* reduced the prevalence of these crimes while the intervention was ongoing. We do, however, find evidence of a statistically significant *increase* in crime after the intervention was complete. Relative to control blocks, on average we observe 0.110 more crimes on treatment blocks and 0.083 more crimes on spillover blocks between November 19 and December 31, 2019. The former effect is statistically significant at the 95% level, the latter very nearly so ($p = 0.059$); together they are weakly jointly significant ($F = 2.66$). They constitute substantively significant increases of 93% and 70%, respectively, relative to the control group mean (0.118 crimes per block after the intervention was complete). Given the number of treatment, spillover, and control blocks in the sample, in aggregate these point estimates imply 23 more crimes in the treatment group (distributed across 214 blocks) and 63 more crimes in the spillover group (distributed across 765 blocks) over the six week period between the end of the intervention and the end of the year. As we show in Appendix H, these adverse effects are driven in particular by an increase in robberies.

Table 1 also reports the effects of the program on standardized additive indices of crime victimization in the endline survey (columns 3 and 4). Respondents were asked if they or anyone in their household had been the victim of any of 10 crimes in the past six months: vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, domestic violence, and extortion. Respondents were also asked the month in which each crime occurred, and, for crimes that occurred in November 2019, whether they occurred before or after the massive nationwide protests that coincided with the end of the intervention. We code dummies for each crime, then construct standardized indices of crime victimization. We find no evidence that *Plan Fortaleza* reduced crime victimization either during or after the intervention. This is inconsistent with the increase in crime in the administrative data, and may reflect the fact that any given randomly selected survey respondent is relatively unlikely to have been a victim of crime in the recent past.[13] We consider this and other possible interpretations in further detail in the

H.

[13]The null effect on crime victimization does not appear to be an artifact of the timing of the endline survey, which was conducted between two and eight weeks after the date of the last crime in the administrative data. Our results are substantively similar if we focus on victims' reports of crimes that occurred between the end of the intervention

conclusion.

Finally, Table 1 reports the effects of the program on a standardized additive index of crime witnessing in the endline survey (column 5). Respondents were asked how frequently they had witnessed each of 15 crimes in the past month. In addition to the 10 crimes listed above, they were asked about prostitution, illegal drug sales, illegal drug consumption, public alcohol consumption, and illegal possession of a firearm. 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 they witnessed must have occurred after *Plan Fortaleza* ended in November 2019. Frequency was measured on a Likert scale from 1 to 4. We use these reports to construct a standardized additive index of crimes witnessed after the intervention.

Consistent with the increase in crime in the administrative data, we find that *Plan Fortaleza* increased reports of witnessing crimes by 0.15 standard deviations on treatment blocks and 0.19 standard deviations on spillover blocks. These are substantively large and highly statistically significant effects, and, as we show in Appendix H, they hold across almost all categories of crime in the index. Also consistent with the increase in crime in the administrative data and the increase in crime witnessing in the endline survey, in exploratory analyses[14] in Tables A.31 and A.32 in the appendix we show that endline survey respondents on both treatment and spillover blocks were more likely to report crimes to the police and more likely to observe police officers establishing a physical presence and making arrests on their blocks. All of these results point to an increase in crime after the intervention was complete. We explore several possible interpretations of this finding in the conclusion.

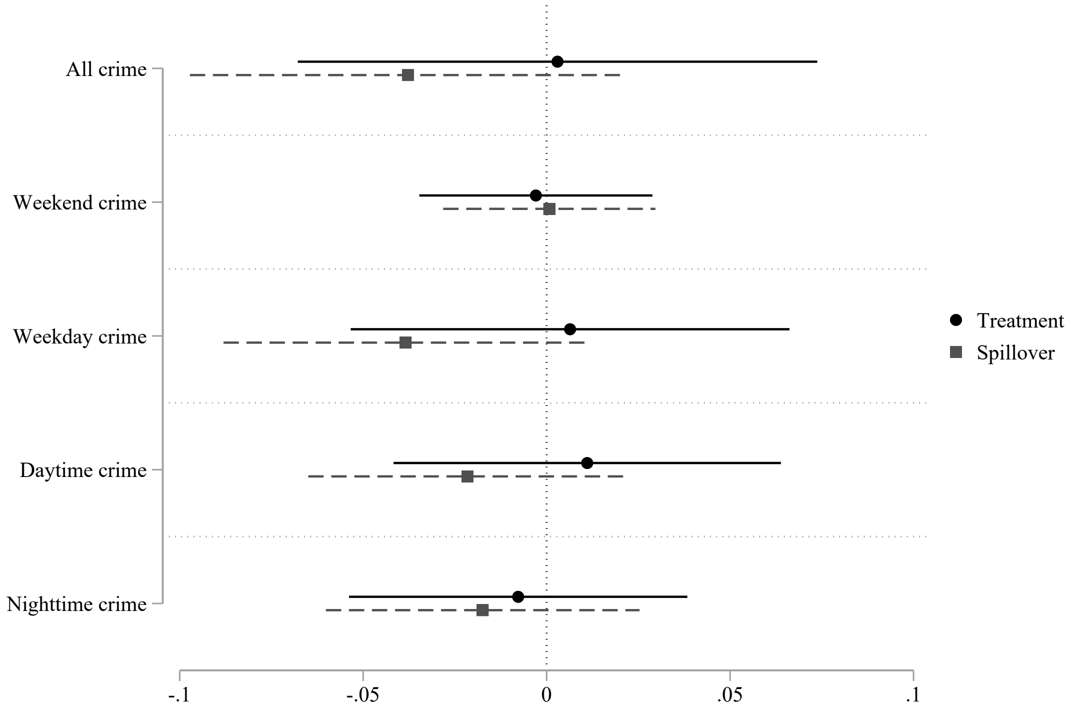
CRIME DISAGGREGATED BY DAY AND TIME

In the aggregate we find no evidence that *Plan Fortaleza* reduced the prevalence of crime while the intervention was ongoing. But these nulls may mask variation by day and time. All *Plan Fortaleza* patrols occurred on weekday nights. Following our PAP, in Figure 2 we use administrative data

and the end of the year.

[14]These analyses were not pre-specified in our PAP, and so should be interpreted as exploratory.

Figure 2: **Treatment effects on crime by day and time**



Notes: ITT on crime during the intervention based on administrative data, disaggregated by day and time. All specifications include neighborhood fixed effects and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Lines denote 95% confidence intervals.

to distinguish crimes committed when soldiers were physically present on the streets (during the week and at night) from those committed when soldiers were physically absent (on the weekend and during the day) while the intervention was ongoing. While we find some suggestive evidence that the program reduced crime while soldiers were physically present on the streets, the effects are relatively small and not statistically significant at conventional levels. Compared to control blocks, we observe 0.038 fewer crimes on spillover blocks on weekdays and 0.017 fewer crimes at night, though neither effect is statistically significant. These point estimates imply reductions of 36% and 23% relative to their respective control group means (0.107 crimes on weekdays and 0.075 crimes at night during the intervention). We again find no evidence of reduced crime on treatment blocks while the intervention was ongoing, even on days and times when soldiers were physically present on the streets.

Table 2: **Treatment effects on perceptions of safety**

	Survey data		
	Perceptions of security		
	All safety	**Personal safety**	**Business safety**
Treatment	-0.050	-0.052	0.284**
	(0.05)	(0.05)	(0.12)
Spillover	-0.066	-0.068*	0.094
	(0.04)	(0.04)	(0.10)
Individual-level controls	✓	✓	✓
Block-level controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Observations	7707	7708	1041
R^2	0.09	0.09	0.13
Control mean	0.077	0.078	-0.065

Notes: ITT on perceptions of safety for all respondents (column 1), residents (column 2), and business owners (column 3) 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.

PERCEPTIONS OF SAFETY

Table 2 reports the effects of the intervention on perceptions of safety in our endline survey. We report results for all respondents together (column 1) and for residents and business owners separately (columns 2 and 3, respectively).[15] Respondents were asked how safe they feel walking their blocks during the day and at night; how safe they feel talking on a smartphone on the street; and how worried they are about becoming victims of 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 streets or attending school. Business owners were also asked if they had closed their businesses, changed their hours, or hired private security guards for fear of crime. We construct standardized additive indices based on responses to these questions.

[15] Given that we did not preregister hypotheses regarding business owners specifically, this analysis is more exploratory.

Our results are mixed. On the one hand, we find no evidence that *Plan Fortaleza* improved perceptions of safety among residents. If anything the opposite appears to be true: the treatment and spillover effects are both negative and similar in magnitude, and the latter effect is (weakly) statistically significant (column 2)—a reduction of 0.068 standard deviations relative to the control group.[16] On the other hand, we find that the program improved perceptions of safety among business owners by 0.284 standard deviations on treatment blocks. One possible explanation for this discrepancy is that business owners were more likely to be physically present during the patrols, and more likely to interact with the patrolling soldiers, who sometimes purchased goods like food or water from local businesses. Another possible explanation is that business owners were especially sensitive to threats posed by gangs (e.g. extortion), and thus particularly receptive to military patrols. These explanations are speculative, however, and overall the effects on perceived safety are weak.

Abuses

Table 3 reports the effects of the intervention on physical and verbal abuses committed by the police (columns 1 and 3) and military (columns 2 and 4) in the monitoring (columns 1 and 2) and endline surveys (columns 3 and 4). Monitoring survey respondents were asked how many times they had seen or heard about physical or verbal abuses committed by police officers or soldiers in the past two weeks; endline survey respondents were asked if they had seen or heard about any abuses in the past month. We construct counts of abuses using the monitoring survey, and dummies using the endline survey. While we measured physical and verbal abuses separately, most respondents who reported one type of abuse also reported the other, and we collapse them for purposes of analysis.

Because the endline survey was conducted in January and February of 2020, we assume that any abuses reported by endline survey respondents must have occurred after the intervention. While the monitoring survey continued for roughly a month after the intervention was complete, we

[16]We do, however, find some evidence of treatment effect heterogeneity among residents. As we show in Table A.24 in the appendix, the program was more likely to improve perceptions of security among residents of blocks with the highest crime rights at baseline.

Table 3: **Treatment effects on abuses**

	Monitoring data		**Survey data**	
	Police abuse	**Military abuse**	**Police abuse**	**Military abuse**
Treatment	0.034***	0.010*	0.011	-0.001
	(0.01)	(0.01)	(0.01)	(0.00)
Spillover	0.011	0.001	0.030**	0.002
	(0.01)	(0.00)	(0.01)	(0.00)
Individual-level controls	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓
Observations	2085	2085	7908	7908
R^2	0.05	0.04	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. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

allow for some degree of telescoping, and assume that any abuses reported by monitoring survey respondents *may* have occurred during the intervention. In Appendix I we plot the distribution of abuses across the two *comunas* in our sample and probe the robustness of our results to different approaches to distinguishing abuses that occurred during the intervention from those that occurred after.

We find some suggestive evidence that *Plan Fortaleza* increased the prevalence of abuses by soldiers on treatment blocks during the intervention, though the effects are substantively small and only weakly statistically significant. Monitoring survey respondents reported 0.01 more incidents of abuse by soldiers on treatment blocks; this effect is only statistically significant at the 90% level. We find no evidence of increased abuses by soldiers on spillover blocks. Importantly, reports of military abuse are exceedingly rare: just 10 out of 2,085 monitoring survey respondents (0.48% of the sample) reported *either* verbal *or* physical abuse by soldiers. In almost all cases, physical and verbal abuses seem to have occurred in the context of the same incidents.[17] Equally importantly,

[17]One respondent reported physical abuse by soldiers but not verbal abuse, and one reported verbal abuse but not

none of the 10 monitoring survey respondents who reported an incident of military abuse was surveyed while the intervention was actually ongoing.[18]

We find more robust evidence of increased physical and verbal abuses by the police. Compared to control blocks, monitoring survey respondents on treatment blocks reported 0.034 more incidents of physical or verbal abuse by police officers. Police abuse was more common than military abuse, with 72 monitoring survey respondents (3.45% of the sample) reporting at least one incident of physical or verbal abuse by a police officer. Again, in most of these cases, physical and verbal abuses appear to have occurred in the context of the same incidents.[19] Roughly half of the monitoring survey respondents who reported police abuse were surveyed while the intervention was ongoing, and roughly half were surveyed after. With just one exception, *all* monitoring survey respondents who reported military abuse reported police abuse as well.

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 physical or verbal 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, military abuse is exceedingly rare in the endline survey: only 65 out of 7,913 endline survey respondents (0.82% of the sample) reported an incident of physical abuse by soldiers, and only 58 (0.73% of the sample) reported an incident of verbal abuse.[20] Police abuse is again much more common: 822 endline survey respondents (10.39% of the sample) reported an incident of physical abuse by police officers, and 740 (9.35% of the sample) reported an incident of verbal abuse.[21]

---

physical abuse. Eight respondents reported both.

[18]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. It is unclear if these abuses were related to *Plan Fortaleza*.

[19]15 respondents reported physical abuse by police officers but not verbal abuse, and 15 reported verbal abuse but not physical abuse. 42 respondents reported both.

[20]There is less overlap between these two groups of respondents in the endline survey. 38 endline survey respondents reported physical abuse but not verbal abuse by soldiers, and 29 reported verbal abuse but not physical abuse. 32 respondents reported both.

[21]Again, there is less overlap between these two groups of respondents in the endline survey. 301 endline survey respondents reported physical abuse but not verbal abuse by police officers, and 216 reported verbal abuse but not physical abuse. 528 respondents reported both.

In contrast, we find little to no evidence of increased abuses in the Attorney-General's data or the detailed firsthand observations of civilian monitors. The monitors recorded only one minor incident of verbal abuse and no incidents of physical abuse at any time during the intervention. Only one allegation in the Attorney-General's data occurred within either of our two study *comunas* during the intervention.[22] This incident occurred on October 29, at the intersection of two spillover blocks and approximately 38 meters from the nearest treatment block in *comuna* 18. The incident involved a transit police officer who was accused of unfairly restricting a citizen's freedom of movement, probably in relation to existing regulations on car and motorcycle traffic in Cali. It seems unlikely that this incident was related to *Plan Fortaleza* in any direct way.

Another two allegations of abuse occurred within our two study *comunas* after the intervention was complete, both involving police officers. Given the nature and timing of the incidents, it seems similarly unlikely that either was related to *Plan Fortaleza*.[23] Taken together, our results suggest that the program increased the prevalence of abuses by police officers, though this finding is somewhat sensitive to data source and specification. To the extent that *Plan Fortaleza* increased the prevalence of military abuses, the effect is very small and also sensitive to data source and specification. We discuss the apparent increase in police abuse in the conclusion.

CONCLUSION

Governments across the developing world increasingly rely on their armed forces for domestic policing operations. These *mano dura* interventions are motivated in part by (legitimate) concerns about persistently high crime rates, and about the ability and willingness of the police to effectively

[22] Two more allegations occurred near but outside of *comuna* 20. They were sufficiently far away from the nearest treatment block—236 meters and 170 meters, respectively—that we believe they were likely unrelated to *Plan Fortaleza*.

[23] The first allegation occurred in *comuna* 18 on the afternoon of November 21 on a spillover block, 173 meters from the nearest treatment block. The witness who reported the incident was walking with her husband when two men ran by. A group of police officers caught the men and—according to the witness—began to beat them, likely in the process of effecting an arrest. When the witness began filming the incident, the officers allegedly attacked her and her husband and seized both of their cell phones, presumably to destroy the evidence. The second allegation occurred on December 6 on a control block in *comuna* 20, 79 meters from the nearest treatment block, and 33 meters from the nearest spillover block. This is one of a series of allegations between the same victim and police officer; in this case the police officer (allegedly) pepper sprayed the victim after the victim (also allegedly) threw rocks at him.

combat crime. Soldiers may be better equipped to credibly threaten and use overwhelming force against organized criminal groups. They may also be less susceptible to corruption and abuse. In much of the Global South, and in Latin America in particular, the public overwhelmingly supports military involvement in law enforcement [11, 39]. Yet these measures have been adopted with little to no evidence regarding their efficacy. We evaluate the effects of an especially common form of military policing—recurring, intensive military patrols targeting crime hot spots—in Cali, Colombia, one of the world's most dangerous cities.

Advocates of military policing argue that soldiers are more effective than police officers at deterring and incapacitating criminals. Our results do not support this claim. To the contrary, we find that military policing exacerbated crime in Cali after the intervention ended. We observe a substantively large and statistically significant increase in crime in the six week period following the end of the intervention. We also observe an increase in respondents' accounts of witnessing and reporting crimes to the police in our endline survey, and of observing police officers making arrests on their blocks. Perhaps relatedly, we find little to no evidence that military policing improved subjective perceptions of safety among residents.

While we do not observe a corresponding increase in crime victimization in the endline survey, the probability that any given respondent was a victim of crime is relatively low.[24] It is possible that crime victimization at endline was not sufficiently common, regardless of treatment assignment, for us to detect a statistically significant treatment effect in either direction. This explanation is speculative, however, and ultimately we cannot be sure. Nonetheless, the preponderance of the evidence from multiple data sources is consistent with an increase in crime in the months after the intervention.

What explains these apparently adverse effects of the *Plan Fortaleza* program? There are several possible explanations. One is that police officers abandoned treatment and spillover blocks after the intervention, perhaps because they assumed (incorrectly) that soldiers would take their place on a more permanent basis. In this case, the adverse effects on crime would not be a result

[24] 12% of endline survey respondents reported being victims of at least one crime in the two to three month period between the end of the intervention and the endline survey.

of military policing per se, but rather of police officers' reactions to the presence of soldiers on their beats. Our results, however, are not consistent with this explanation. If anything we observe *greater* police presence and activity on treatment and spillover blocks after the intervention was complete. Negligence by police officers cannot explain these patterns in the data.

A second possible explanation is that increased crime in the administrative data is an artifact of increased crime witnessing and reporting in the endline survey. In Cali, as in most cities, many crimes go unreported; it is possible that military policing induced heightened vigilance among civilians (perhaps because they interpreted the military's presence as a signal that their blocks were less safe than they previously believed), and that the crimes they witnessed and reported were subsequently entered into the city's administrative records when police officers responded to their complaints. This would help explain why we observe an increase in crime in the administrative data without a corresponding increase in crime victimization in the endline survey. It would also suggest that the increase in crime in the administrative data is an illusion, and masks a more benign—perhaps even beneficial—increase in cooperation between civilians and the police.

While we cannot rule out this explanation conclusively, it is inconsistent with some of our other results. If increased crime witnessing and reporting in the endline survey explained increased crime in the administrative data, we would expect to find a close correspondence between the types of crimes that respondents witnessed in the endline survey and the types of crimes that appeared in the city's administrative records. But we do not. Relative to control blocks, endline survey respondents were much more likely to witness homicides, drug dealing, and illegal possession of firearms on treatment and spillover blocks. But we find no evidence of an increase in homicides or drug dealing in the administrative data, and only weak (and not statistically significant) evidence of an increase in illegal possession of firearms. Heightened vigilance might explain the positive effect on witnessing these crimes in the survey, but it cannot explain the null effect on these same crimes in the administrative data.

A third possible explanation is that criminals were not (or were only slightly) deterred by the military's presence during the intervention, but were nonetheless emboldened by the military's absence after the intervention was complete. In other words, there may be negligible benefits to deploying soldiers to crime hot spots, but serious consequences to withdrawing them. This would distinguish military policing from hot spots policing, which generally deters crime both during and after implementation, with minimal crime displacement and some diffusion of crime control benefits [13, 14]. Few studies directly assess the consequences of *withdrawing* police presence; the existing evidence is limited and contradictory [40, 41]. While analogies to war zones and insurgencies should be drawn cautiously, recent research suggests that the withdrawal of US armed forces from Afghanistan precipitated a rise in violence [42], though evidence on the effects of drawing down military presence is again limited. Whatever the explanation, the apparent increase in crime after *Plan Fortaleza* was complete belies one of the most significant purported benefits of military policing.

Critics of military involvement in domestic policing operations often argue that soldiers are more likely than police officers to abuse civilians. Our results suggest this dynamic may be more complex than critics contend. On the one hand, we find no evidence of increased abuses in administrative data from the Attorney-General's Office, or in the firsthand observations of civilian monitors. On the other hand, we do find evidence of increased abuses in surveys administered to residents both during and (weakly) after the intervention. But in most cases these abuses appear to have been perpetrated by police officers rather than soldiers. Under Colombian law, the military can interrogate and detain suspects, but only the police can make arrests. As we show in Table A.31 in the appendix, residents were more likely to observe police officers making arrests on both treatment and spillover blocks, both during and after the intervention, suggesting some degree of "co-production" of security between the military and the police. In this way, military policing in Cali may have created additional opportunities for police officers to commit abuse.

Taken together, our findings suggest that the costs of military policing outweigh the benefits. While Cali's *Plan Fortaleza* program had complex and in some cases counterintuitive effects on both objective and subjective measures of security, on balance our results do not support the claims of proponents that military policing is necessary to deter crime and mitigate abuses by predatory police forces. Military policing interventions like *Plan Fortaleza* have become increasingly pervasive and increasingly popular in recent years, especially in Latin America. Our evaluation, combined with the similarly disappointing results of several recent observational analyses [5, 25, 26], suggest that this trend is unlikely to improve security in the world's most dangerous cities. Policymakers should consider alternative crime reduction strategies that are less likely to produce unintended adverse effects.

Data availability statement

All data generated or analysed during this study will be included with this published article (and its supplementary information files).

Code availability statement

All code required to replicate the analyses in this study will be posted publicly upon publication.

REFERENCES

1. Giddens, A. *The Nation-State and Violence* (University of California Press, Berkeley, 1987).
2. Huntington, S. P. *The Soldier and the State: The Theory and Politics of Civil–Military Relations* (Belknap, Cambridge, MA, 1957).
3. Kraska, P. B. Militarization and Policing–Its Relevance to 21st Century Police. *Policing* **1,** 501–513 (2007).
4. Collazos, D., Garcia, E., Mejia, D., Ortega, D. & Tobón, S. *Hot Spots Policing in a High Crime Environment: An Experimental Evaluation in Medellín* SSRN Scholarly Paper 3316968. 2019.
5. Flores-Macías, G. & Zarkin, J. *The Consequences of Militarized Policing for Human Rights: Evidence from Mexico* Presented at the International Studies Association Annual Conference, April 6–9. 2021.
6. Meliala, A. Police as military: Indonesia's experience. *Policing: An International Journal of Police Strategies & Management* (2001).
7. Varona, G. Politics and policing in the Philippines: challenges to police reform (2010).
8. Montesh, M. & Basdeo, V. The role of the South African national defence force in policing. *Scientia Militaria: South African Journal of Military Studies* **40** (2012).
9. Dunlap Jr., C. J. The Police-Ization of the Military. *Journal of Political and Military Sociology* **27,** 397–418 (1999).
10. Holland, A. C. Right on Crime?: Conservative Party Politics and Mano Dura Policies in El Salvador. *Latin American Research Review* **48,** 44–67 (2013).
11. Pion-Berlin, D. & Carreras, M. Armed Forces, Police and Crime-Fighting in Latin America. *Journal of Politics in Latin America* **9,** 3–26 (2017).

12. Muggah, R., Garzón, J. C. & Suárez, M. *Mano Dura: The costs and benefits of repressive criminal justice for young people in Latin America* (Igarapé Institute, Rio de Janeiro, 2018).

13. Braga, A., Papachristos, A. & Hureau, D. Hot Spots Policing Effects on Crime. *Campbell Systematic Reviews* **8,** 1–96 (2012).

14. Braga, A. A., Turchan, B. S., Papachristos, A. V. & Hureau, D. M. Hot spots policing and crime reduction: an update of an ongoing systematic review and meta-analysis. *Journal of Experimental Criminology* **15,** 289–311 (2019).

15. Bove, V. & Gavrilova, E. Police Officer on the Frontline or a Soldier? The Effect of Police Militarization on Crime. *American Economic Journal: Economic Policy* **9,** 1–18 (2017).

16. Delehanty, C., Mewhirter, J., Welch, R. & Wilks, J. Militarization and police violence: The case of the 1033 program. *Research & politics* **4,** 1–7 (2017).

17. Gunderson, A. *et al.* Counterevidence of Crime-Reduction Effects from Federal Grants of Military Equipment to Local Police. *Nature Human Behaviour* **5,** 194–204 (2021).

18. Harris, M. C., Park, J., Bruce, D. J. & Murray, M. N. Peacekeeping Force: Effects of Providing Tactical Equipment to Local Law Enforcement. *American Economic Journal: Economic Policy* **9,** 291–313 (2017).

19. Lowande, K. Police Demilitarization and Violent Crime. *Nature Human Behaviour* **5,** 205–211 (2021).

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

21. González, Y. M. *Authoritarian Police in Democracy: Contested Security in Latin America* (Cambridge University Press, Cambridge, UK, 2020).

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

23. Magaloni, B., Franco-Vivanco, E. & Melo, V. Killing in the Slums: Social Order, Criminal Governance, and Police Violence in Rio de Janeiro. *American Political Science Review* **114,** 552–572 (2020).

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

25. Espinosa, V. & Rubin, D. B. Did the Military Interventions in the Mexican Drug War Increase Violence? *The American Statistician* **69,** 17–27 (2015).

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

27. Blattman, C., Green, D., Ortega, D. & Tobón, S. *Place-Based Interventions at Scale: The Direct and Spillover Effects of Policing and City Services on Crime* NBER Working Paper No. 23941. 2019.

28. Collazos, D., Garcıa, E., Mejıa, D., Ortega, D. & Tobón, S. *Hot Spots Policing in a High Crime Environment: An Experimental Evaluation in Medellın* SSRN Scholarly Paper 3316968. 2019.

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

30. den Heyer, G. Mayberry Revisited: A Review of the Influence of Police Paramilitary Units on Policing. *Policing and Society* **24,** 346–361 (2014).

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

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

33. Tecott, R. & Plana, S. Maybe U.S. Police Aren't Militarized Enough. Here's What Police Can Learn from Soldiers. *Monkey Cage* **August 16** (2016).

34. Bayley, D. H. Criminal Investigation: Introduction. *What Works in Policing,* 71–74 (1998).

35. Lessing, B. Logics of Violence in Criminal War. *Journal of Conflict Resolution* **59,** 1486–1516 (2015).

36. Flores-Macías, G. A. & Zarkin, J. The Militarization of Law Enforcement: Evidence from Latin America. *Perspectives on Politics* **19,** 519–538 (2021).

37. Aronow, P. M. & Samii, C. Estimating Average Causal Effects under General Interference, with Application to a Social Network Experiment. *The Annals of Applied Statistics* **11,** 1912–1947 (2017).

38. Bowers, K. J., Johnson, S. D., Guerette, R. T., Summers, L. & Poynton, S. Spatial displacement and diffusion of benefits among geographically focused policing initiatives: a meta-analytical review. *Journal of Experimental Criminology* **7,** 347–374 (2011).

39. Flores-Macıas, G. & Zarkin, J. *Militarization and Perceptions of Law Enforcement in the Developing World: Evidence from a Conjoint Experiment in Mexico* Presented at the American Political Science Association Annual Meeting, Washington, DC, August 28–September 1. 2019.

40. DeAngelo, G. & Hansen, B. Life and Death in the Fast Lane: Police Enforcement and Traffic Fatalities. *American Economic Journal: Economic Policy* **6,** 231–257 (2014).

41. Sullivan, C. M. & O'Keeffe, Z. P. Evidence That Curtailing Proactive Policing Can Reduce Major Crime. *Nature Human Behaviour* **1,** 730–737 (2017).

42. Fetzer, T., Eynde, O. V., Wright, A. L. & Souza, P. C. *Security Transitions* Working paper. 2019.

43. Anderson, M. L. 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,** 1481–1495 (2008).

Appendix

Appendix

A Setting and implementation

Table A.1 illustrates the 12-day rotation schedule used by the Special Forces and Military Police during the intervention. The purpose of the schedule was to avoid logistical problems by ensuring that the two units never patrolled the same *comuna* on the same day. Figures A.1 and A.2 show screenshots from the Google Maps platform used by civilian monitors to help the soldiers navigate to treatment blocks and record the soldiers' activities. Table A.2 provides descriptive statistics on the soldiers' activities while on patrol based on the monitors' reports. Figure A.5 plots the distribution of homicides across the 22 *comunas* in Cali based on administrative data for 2019.

B Randomization

We stratified by neighborhood, then randomized such that approximately 1/6 of all blocks in each neighborhood were assigned to treatment. We assigned to the spillover group any block that was adjacent to at least one treatment block, but was not itself assigned to treatment. 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. Figures A.3 and A.4 map the blocks in our sample by treatment assignment.

C Power calculations

We used administrative crime data from the first six months of 2019 to simulate minimum detectable effects (MDEs). We first bootstraped different randomization procedures 1,500 times each. Within each replication we simulated the intention-to-treat effect (ITT), knowing that the true ITT was 0. For simplicity we focused on estimating the ITT on an unweighted additive index of homicides and robberies on each block in our sample. We focused on these two crimes because they were the only ones for which we had complete, georeferenced administrative data before registering our pre-analysis plan. We also simulated spillover effects, knowing that the true spillover effect was 0 as well.

We estimated the ITT using a weighted least squares regression given by

$$y_{jk} = \theta t_{jk} + \lambda s_{jk} + \beta \mathbf{X_{jk}} + \alpha_k + \epsilon_{jk}$$

where y_{jk} denotes crime on block j in neighborhood k; t_{jk} denotes assignment to treatment; s_{jk} denotes assignment to spillover; α_k denotes neighborhood fixed effects; and ϵ_j 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 calculated the standard deviation of these simulated ITTs. To estimate our MDEs, we simply multiplied the standard deviation of the simulated effects by 2.49 for a two-tailed test. We did not adjust for covariates when simulating MDEs. We repeated this exercise for different ratios of treatment to control and treatment to spillover blocks. When assigning 1/6 of all blocks in each neighborhood to treatment, we are powered to detect ITTs of approximately 0.25 standard deviations relative to the control group. We are powered to detect spillover effects of approximately 0.20 standard deviations when comparing our spillover group to control. We would have gained little to no power for estimating treatment or spillover effects by increasing the proportion of blocks assigned to treatment above 1/6.

D Balance tests

Tables A.3 and A.4 report balance tests using our administrative crime data and endline survey data, respectively.

E SPILLOVER

In the paper we assume that spillover can only occur from treated blocks to adjacent untreated blocks. Here we relax this assumption in three ways. First, we assume that treatment effects become weaker as a linear function of distance to the nearest treated block. We estimate

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

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 treated blocks, where *t* indicates treated blocks. We calculate the distance, *d*, to all treated blocks; $\frac{1}{d}$ for each block; sum the distances to all treated 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 treated block.

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

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

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 treated blocks; $\frac{1}{e^d}$ for each block; sum the distances to all treated 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 treated block.

Finally, we assume that the strength of the treatment effect on any given treated 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. All analyses include block-level controls and neighborhood fixed effects; all estimates are weighted by the inverse probability of assignment to each block's realized treatment status.

As we show in Table A.5, we find little evidence of these more complex spillover dynamics at work. The coefficients on the linear and exponential decay variables are negative when we focus on crimes committed after the intervention was complete, but they are only weakly statistically significant at conventional levels.

F 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 use the weighted index in our main specifications; in the paper we instead use the unweighted index in our main specifications, and report results using the weighted index in Section H below. Based on feedback from criminologists, we determined that our approach to weighting is not standard in the literature, and yields results that are difficult to interpret substantively.

In our PAP we proposed to collect administrative crime data on homicide, assault, theft, car theft, and motorcycle theft. We were in fact able to collect administrative crime data on homicide, robbery (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 of these crimes in our index. In our PAP we also proposed to test the ITT of the program on arrests based on administrative data. Unfortunately we were unable to obtain data on arrests from the government of Cali, so we drop this analysis here.

In our PAP we proposed to compute Lee bounds to estimate 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. We also proposed to estimate more complex spillover dynamics using a marginalized individualistic response function, following **?**. We decided to drop this analysis because the procedure is relatively new and untested. Finally, in our PAP we posited several additional hypotheses related to perceptions of the police and military, political beliefs, and voting behavior. For compactness we focus in this paper on crime, perceptions of safety, and abuses, as these are the outcomes of most urgent concern to both proponents and detractors of military policing. We report treatment effects on perceptions, political beliefs, and voting behavior in a separate paper.

G Multiple comparisons

In the paper we use indexing to reduce the number of hypotheses we test. Here we replicate all analyses and report the Benjamini-Hochberg q-value and Holm-Bonferroni threshold for our treatment and spillover indicators. Following [43], 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 $\alpha = 0.05$.

We apply each correction within (but not across) "families" of outcome. Tables A.6 and A.7, for example, amount to a test of the hypothesis that *Plan Fortaleza* affected crime at all, across all of our various proxies for crime. For each table we produce a "stringent" and "lenient" version of the multiple comparisons correction: the "stringent" version assumes that for each model we test *two* hypotheses—one corresponding to treatment effects and the other to spillover effects—while the "lenient" version assumes that for each model we test one combined hypothesis that the program had any effect on the outcome.

H Crime

H.1 Weighted standardized crime indices

Table A.12 reports the ITT of the *Plan Fortaleza* program on weighted standardized indices of crimes committed during (column 1) and after (column 2) the intervention using administrative data. Crimes are weighted by the average prison sentence under Colombian law. Figure A.6

distinguishes crimes committed when soldiers were physically present on the streets (during the week and at night) from those committed when soldiers were physically absent (on the weekend and during the day) during the months when the intervention was ongoing.

H.2 Violent and non-violent crime indices

Table A.13 reports the ITT of the *Plan Fortaleza* program on indices of violent and non-violent crimes committed during (columns 1 and 2) and after (columns 3 and 4) the intervention using administrative data. Table A.14 reports the ITT on standardized indices of violent crime victimization and witnessing in the survey; Table A.15 reports the ITT on standardized indices of non-violent crime victimization and witnessing.

H.3 Crime disaggregated by type

Tables A.16 reports the ITT of the *Plan Fortaleza* program on specific categories of crime in the administrative data. Tables A.17 and A.18 report the ITT on specific catetogries of crime victimization and witnessing, respectively, in the survey.

I Abuses

I.1 Abuses in administrative data

Figure A.7 shows the distribution of alleged abuses committed by soldiers and police officers across the city of Cali, as reported by victims and witnesses to the Office of the Attorney-General. Red markers denote allegations from the period during the intervention, and blue markers denote allegations from the period after. In some cases complaints were lodged days or weeks after the alleged abuse occurred. In these cases we distinguish the period during the intervention from the period after using the date the abuse was alleged to have occurred, rather than the date the report was filed. Figures A.8 and A.9 focus on the two *comunas* in our sample. Green denotes treatment blocks, yellow spillover blocks, and red control blocks.

I.2 Abuses in monitoring survey data

Figures A.10 through A.17 plot the distribution of abuses committed by soldiers and police officers across the two *comunas* in our sample, as reported by monitoring survey respondents. While the monitoring survey continued for roughly one month after the intervention was complete, we allow for some degree of telescoping, and assume that any abuses reported by monitoring survey respondents may have occurred during the intervention, regardless of when those respondents were surveyed.

I.3 Abuses in monitoring survey disaggregated by period

Tables A.19 and A.20 replicate our results in columns 1 and 2 of Table 3 using two different approaches to defining the period during and after the intervention. In Table A.19 we assume that any abuses reported by respondents who were surveyed before November 19, 2019 (the day 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 on abuses that they witnessed or heard about in the two weeks prior to being surveyed. As an additional robustness check, in Table A.20 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.

J Heterogeneous treatment effects

Table A.21 reports heterogeneous treatment effects on crime in the administrative data and crime victimization and witnessing in the endline survey by prior crime rate. To operationalize prior crime rate, we construct an additive index of crimes committed on each block between January 1, 2019 and the start of the intervention on September 30, 2019 based on administrative data. Table A.23 reports heterogeneous treatment effects on crime victimization and witnessing by gender. (Since crime is operationalized at the block level and gender is operationalized at the individual level, we do not test for heterogeneous treatment effects on crime in the administrative data by gender in the survey.) Tables A.24 and A.26 report heterogeneous treatment effects on perceptions of security by prior crime rate in the administrative data and gender, respectively. Tables A.27 and A.29 report heterogeneous treatment effects on abuses by prior crime rate and gender, respectively.

Table A.22 reports heterogeneous treatment effects on crime victimization and witnessing by prior crime victimization. To operationalize prior crime victimization, we create a standardized additive index of respondents' reports of any crimes committed against them or their family members in August or September 2019 based on the endline survey. Unfortunately we did not ask about crimes committed prior to August 2019. (Since crime is operationalized at the block level and prior crime victimization is operationalized at the individual level, we do not test for heterogeneous treatment effects on crime in the administrative data by prior crime victimization in the survey.)

Table A.25 reports heterogeneous treatment effects on perceptions of safety by prior crime victimization, while Table A.28 reports heterogeneous treatment effects on abuses by prior crime victimization. We use the same operationalization for crime victimization described above.

K ANCILLARY AND EXPLORATORY ANALYSES

K.1 EXPOSURE TO POLICE AND MILITARY

Table A.30 reports the effect of the *Plan Fortaleza* program on exposure to the military and the police in the monitoring and endline surveys. We asked monitoring survey respondents how often they had seen or heard about the police or military on their block in the prior two weeks (columns 1 and 2). Frequency was measured on a Likert scale from 1 to 4, then standardized. We asked the same question at endline, although we extended the temporal window from two weeks to one month (columns 3 and 4). This question does not measure treatment compliance per se, since endline survey respondents were surveyed a month or more after the program ended. Testing the ITT on this measure helps us assess whether soldiers continued patrolling the same treatment blocks even after the intervention.

K.2 Police and military activities

Table A.31 reports the ITT of the program on military and police activities. Respondents were asked how often they had seen or heard about the police or military making arrests on their block during (columns 1 and 2) and after (columns 7 and 8) the intervention; how often they had seen or heard about the police (columns 3 and 4) or military (column 9 and 10) checking IDs on their block; and how often they had seen or heard about the police or military talking with citizens on their block during (columns 5 and 6) and after (columns 11 and 12) the intervention. The monitoring survey (columns 1-6) asked respondents whether they had seen or heard about these activities in the prior two weeks, while the endline survey (columns 7-12) asked respondents whether they had seen or heard about these activities in the prior month. Frequency was measured on a Likert scale from 1 to 4.

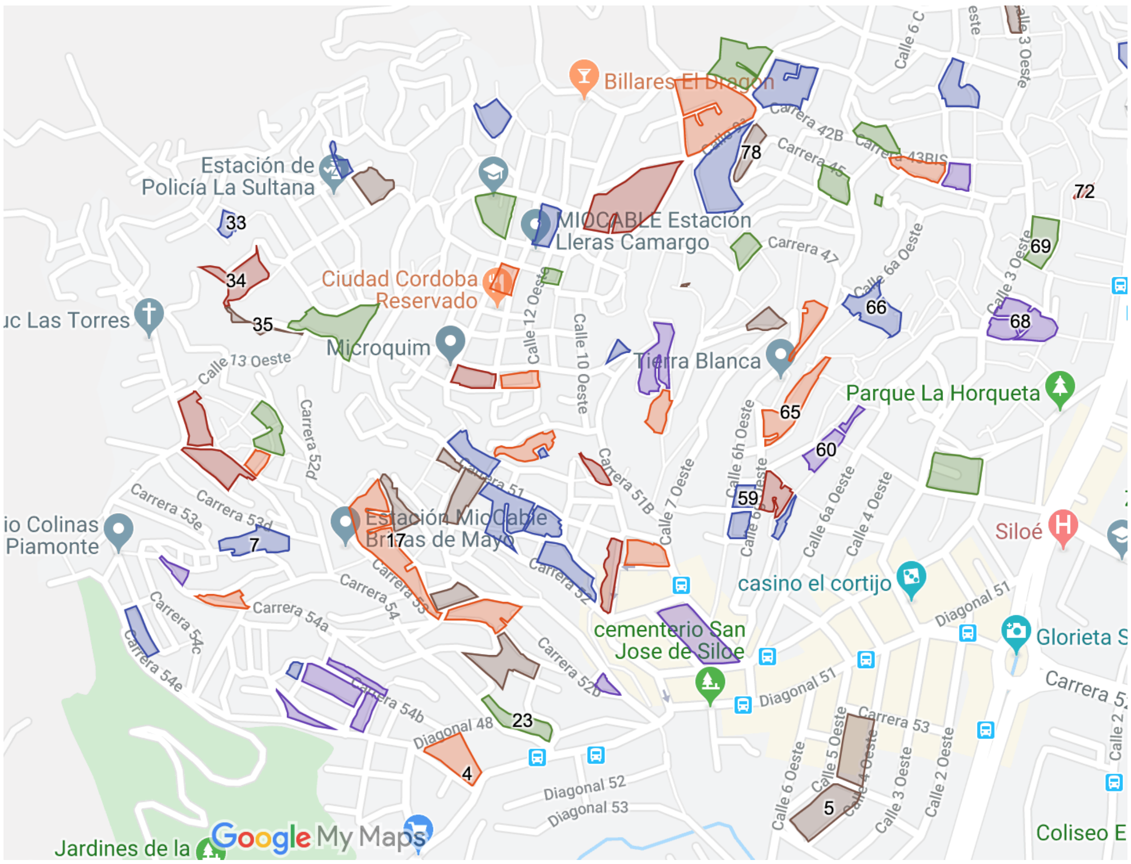
K.3 Cooperation with police and military

Table A.32 reports the effect of the program on cooperation with the authorities in general (column 1), and the police (column 2) and military (column 3) specifically, based on 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 indices of cooperation using responses to these questions.

Table A.1: **12-day rotation schedule for Special Forces and Military Police**

	Comuna 18	*Comuna* 20
Day 1	Special Forces	Military Police
Day 2	Military Police	Special Forces
Day 3	Special Forces	Military Police
Day 4	Military Police	Special Forces
Day 5	Special Forces	Military Police
Day 6	Military Police	Special Forces
Day 7	Military Police	Special Forces
Day 8	Special Forces	Military Police
Day 9	Military Police	Special Forces
Day 10	Special Forces	Military Police
Day 11	Military Police	Special Forces
Day 12	Special Forces	Military Police

Figure A.1: **Screenshot of smartphone user interface with locations of treatment blocks in *comuna* 20**



Notes: Colors indicate which day a treatment block was assigned to be patrolled.

Figure A.2: **Screenshot of smartphone user interface with information on a single treatment block**

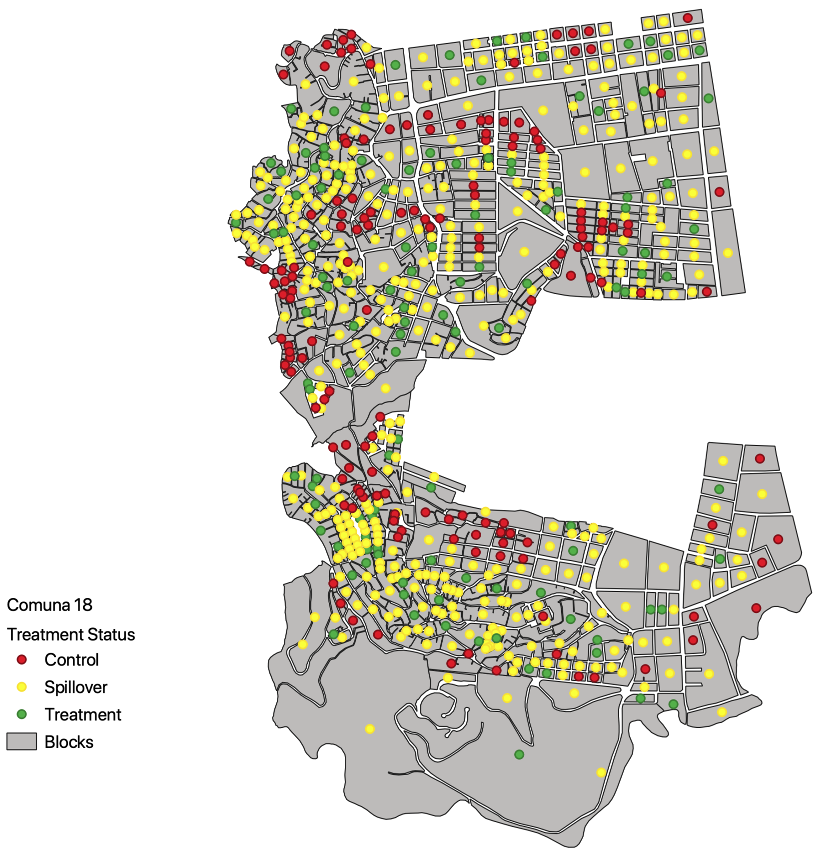
Notes: Information provided to soldiers included latitude and longitude, patrol day, and cross streets for each treatment block.

Table A.2: **Descriptive statistics on patrols**

	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

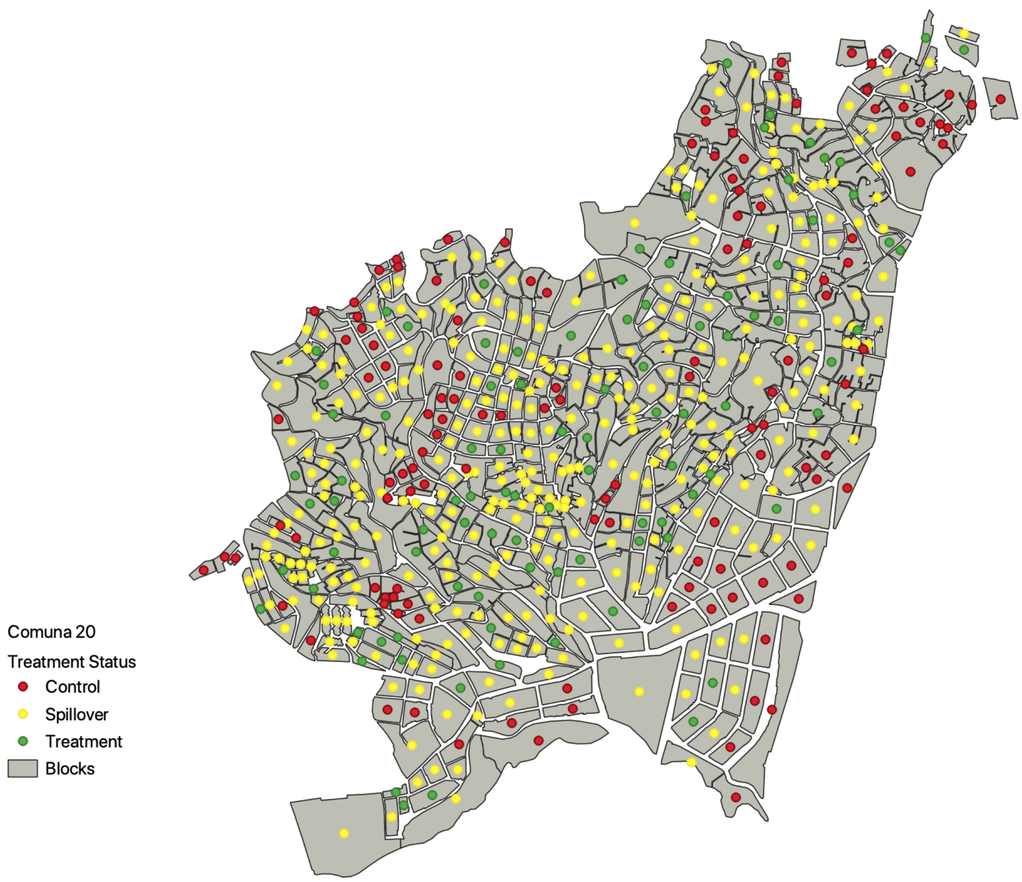
Notes: Standard deviations in parentheses.

Figure A.3: **Assignment to treatment in *comuna* 18**



Notes: Green denotes treatment blocks, yellow denotes spillover blocks, and red denotes control blocks.

Figure A.4: **Assignment to treatment in *comuna* 20**



Notes: Green denotes treatment blocks, yellow denotes spillover blocks, and red denotes control blocks.

Figure A.5: **Homicides in Cali by *comuna* using administrative data**

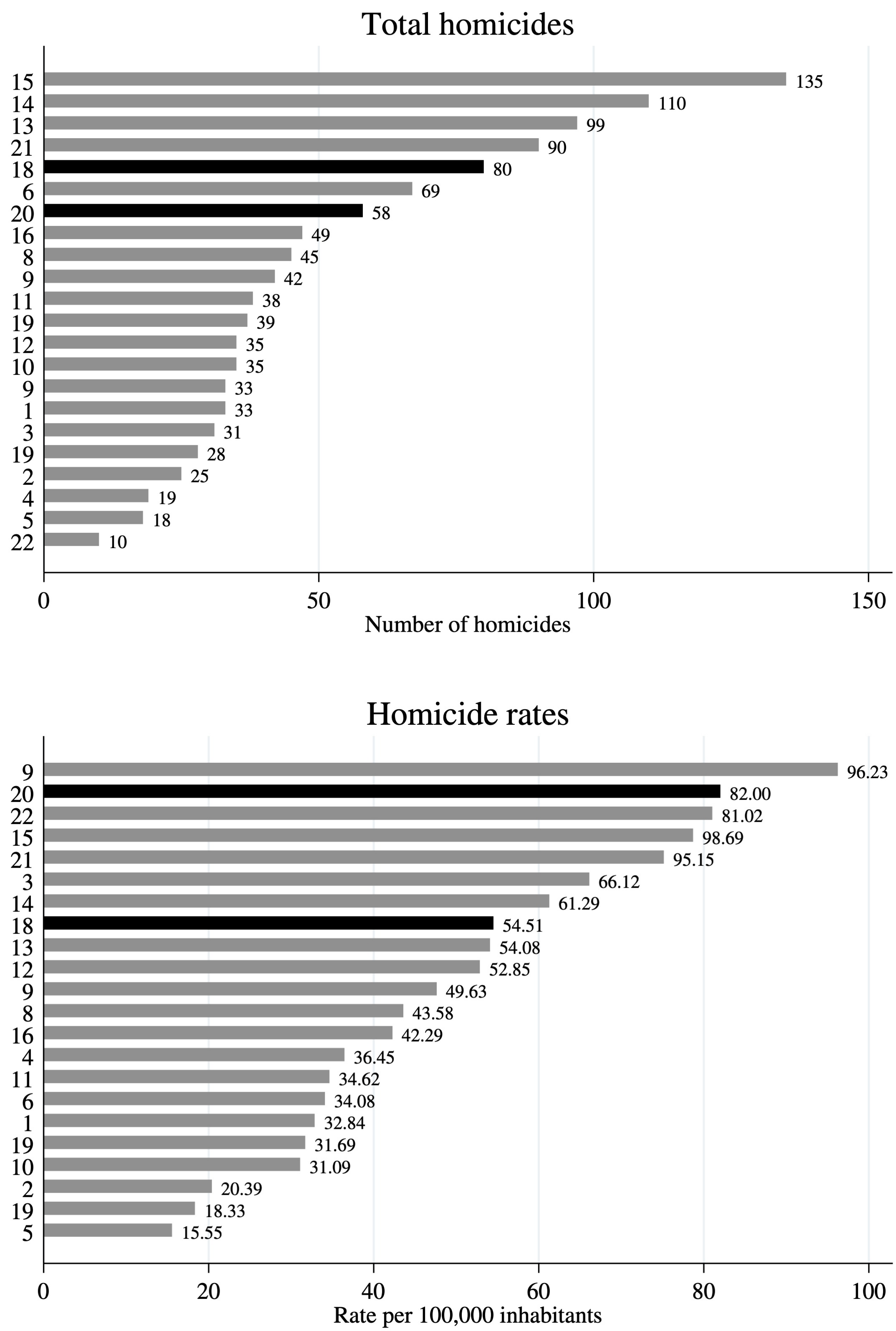


Table A.3: **Balance using administrative data**

	Control	Treatment	Spillover
Panel A: Index without controls			
Crime index	0.002	0.000	-0.002
	(0.007)	(0.010)	(0.004)
Panel B: Components without controls			
Homicides	-0.019	-0.002	0.021
	(0.057)	(0.066)	(0.042)
Robberies	0.004	-0.001	-0.003
	(0.009)	(0.011)	(0.004)
Drug dealing	-0.027	0.077	-0.049
	(0.084)	(0.100)	(0.061)
llegal possession of a firearm	-0.086	0.024	0.062
	(0.073)	(0.126)	(0.095)
Panel C: Index with controls			
Crime index	0.002	0.001	-0.004
	(0.008)	(0.010)	(0.004)
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	-0.023	0.022	0.001
	(0.059)	(0.068)	(0.044)
Robberies	0.004	-0.000	-0.004
	(0.009)	(0.012)	(0.004)
Drug dealing	-0.009	0.060	-0.051
	(0.088)	(0.104)	(0.066)
llegal possession of a firearm	-0.089	0.022	0.067
	(0.070)	(0.122)	(0.098)
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	0.097	0.000	0.276
Panel B: F-stat on components without controls	0.517	0.164	0.433
Panel C: F-stat on index	0.088	0.014	0.756
Panel C: F-stat on index with controls	1.941*	1.218	1.561
Panel D: F-stat on components	0.545	0.128	0.468
Panel D: F-stat on components with controls	1.633	0.868	1.152

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

Table A.4: **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.5: **Treatment effects on crime using alternative approaches to estimating spillover effects**

	Crime index					
	Linear decay		**Exponential decay**		**% 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
	(0.03)	(0.04)	(0.03)	(0.04)	(0.19)	(0.31)
Spillover	-0.007	-0.070*	-0.000	-0.096*		
	(0.02)	(0.03)	(0.04)	(0.05)		
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 crime during (columns 1, 3, and 5) and after (columns 2, 4, and 6) the intervention based on administrative data, using alternative measures to estimate spillover. Columns 1 and 2 use a linear decay function, columns 3 and 4 use an exponential decay function, and columns 5 and 6 use the proportion of adjacent treatment blocks 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.6: **Treatment effects on crime, crime victimization, and witnessing crime using lenient correction for multiple comparisons**

	Admin data		Survey data		
	Crime incidence		Crime victimization		Crime witnessing
	During intervention	After intervention	During intervention	After intervention	After intervention
Treatment	0.003	0.110**	0.006	-0.007	0.153***
	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)
Spillover	-0.038	0.083*	0.026	0.013	0.186***
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)
Individual controls	✗	✗	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓
Observations	1167	1167	7845	7845	7837
R^2	0.33	0.48	0.03	0.03	0.12
Control mean	0.160	0.160	-0.021	-0.016	-0.119
qval-treatment	0.935	0.073	0.935	0.935	0.017
qval-spillover	0.355	0.147	0.487	0.729	0.001
BH-treatment		✓			✓
BH-spillover					✓
Holm-treatment					✓
Holm-spillover					✓

Notes: ITT on crime during (column 1) and after (column 2) the intervention based on administrative data; crime victimization during (column 3) and after (column 4) the intervention based on survey data; and crime witnessing after the intervention (column 5) based on survey data. Check marks indicate whether the corresponding coefficient remains statistically significant after applying the Benjamini-Hochberg or Holm-Bonferroni correction under the lenient assumption that each regression tests one hypothesis. All specifications include neighborhood fixed effects and block-level controls. Models 3-6 also include individual 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.7: **Treatment effects on crime, crime victimization, and witnessing crime using stringent correction for multiple comparisons**

	Admin data		**Survey data**		
	Crime incidence		**Crime victimization**		**Crime witnessing**
	During intervention	After intervention	During intervention	After intervention	After intervention
Treatment	0.003	0.110**	0.006	-0.007	0.153***
	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)
Spillover	-0.038	0.083*	0.026	0.013	0.186***
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)
Individual controls	✗	✗	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓
Observations	1167	1167	7845	7845	7837
R^2	0.33	0.48	0.03	0.03	0.12
Control mean	0.160	0.160	-0.021	-0.016	-0.119
qval-treatment	0.935	0.098	0.935	0.935	0.017
qval-spillover	0.425	0.147	0.649	0.935	0.001
BH-treatment		✓			✓
BH-spillover					✓
Holm-treatment					✓
Holm-spillover					✓

Notes: ITT on crime during (column 1) and after (column 2) the intervention based on administrative data; crime victimization during (column 3) and after (column 4) the intervention based on survey data; crime witnessing after the intervention (column 5) based on survey data. Check marks indicate whether the corresponding coefficient remains statistically significant after applying the Benjamini-Hochberg or Holm-Bonferroni correction under the stringent assumption that each regression tests two hypotheses. All specifications include neighborhood fixed effects and block-level controls. Models 3-6 also include individual 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.8: **Treatment effects on perceptions of security using lenient correction for multiple comparisons**

	Survey data		
	Perceptions of security		
	All safety	**Personal safety**	**Business safety**
Treatment	-0.050	-0.052	0.284**
	(0.05)	(0.05)	(0.12)
Spillover	-0.066	-0.068*	0.094
	(0.04)	(0.04)	(0.10)
Individual-level controls	✓	✓	✓
Block-level controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Observations	7707	7708	1041
R^2	0.09	0.09	0.13
Control mean	0.077	0.078	-0.065
qval-treatment	0.292	0.292	0.045
qval-spillover	0.151	0.151	0.368
BH-treatment			✓
BH-spillover			
Holm-treatment			✓
Holm-spillover			

Notes: ITT on perceptions of safety for all respondents (column 1), residents (column 2), and business owners (column 3) based on survey data. Check marks indicate whether the corresponding coefficient remains statistically significant after applying the Benjamini-Hochberg or Holm-Bonferroni correction under the lenient assumption that each regression tests one hypothesis. 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.9: **Treatment effects on perceptions of security using stringent correction for multiple comparisons**

	Survey data		
	Perceptions of security		
	All safety	**Personal safety**	**Business safety**
Treatment	-0.050	-0.052	0.284**
	(0.05)	(0.05)	(0.12)
Spillover	-0.066	-0.068*	0.094
	(0.04)	(0.04)	(0.10)
Individual-level controls	✓	✓	✓
Block-level controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Observations	7707	7708	1041
R^2	0.09	0.09	0.13
Control mean	0.077	0.078	-0.065
qval-treatment	0.351	0.351	0.089
qval-spillover	0.201	0.201	0.368
BH-treatment			✓
BH-spillover			
Holm-treatment			✓
Holm-spillover			

Notes: ITT on perceptions of safety for all respondents (column 1), residents (column 2), and business owners (column 3) based on survey data. Check marks indicate whether the corresponding coefficient remains statistically significant after applying the Benjamini-Hochberg or Holm-Bonferroni correction under the stringent assumption that each regression tests two hypotheses. 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.10: **Treatment effects on abuses using lenient correction for multiple comparisons**

	Monitoring data		**Survey data**	
	Police abuse	**Military abuse**	**Police abuse**	**Military abuse**
Treatment	0.034***	0.010*	0.011	-0.001
	(0.01)	(0.01)	(0.01)	(0.00)
Spillover	0.011	0.001	0.030**	0.002
	(0.01)	(0.00)	(0.01)	(0.00)
Individual-level controls	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓
Observations	2085	2085	7908	7908
R^2	0.05	0.04	0.05	0.01
Control mean	0.015	0.000	0.114	0.012
qval-treatment	0.007	0.128	0.556	0.716
qval-spillover	0.584	0.650	0.044	0.650
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). Check marks indicate whether the corresponding coefficient remains statistically significant after applying the Benjamini-Hochberg or Holm-Bonferroni correction under the lenient assumption that each regression tests one hypothesis. 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 effects on abuses using stringent correction for multiple comparisons**

	Monitoring data		**Survey data**	
	Police abuse	**Military abuse**	**Police abuse**	**Military abuse**
Treatment	0.034***	0.010*	0.011	-0.001
	(0.01)	(0.01)	(0.01)	(0.00)
Spillover	0.011	0.001	0.030**	0.002
	(0.01)	(0.00)	(0.01)	(0.00)
Individual-level controls	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓
Observations	2085	2085	7908	7908
R^2	0.05	0.04	0.05	0.01
Control mean	0.015	0.000	0.114	0.012
qval-treatment	0.014	0.170	0.668	0.716
qval-spillover	0.584	0.716	0.044	0.716
BH-treatment	✓			
BH-spillover			✓	
Holm-treatment	✓			
Holm-spillover			✓	

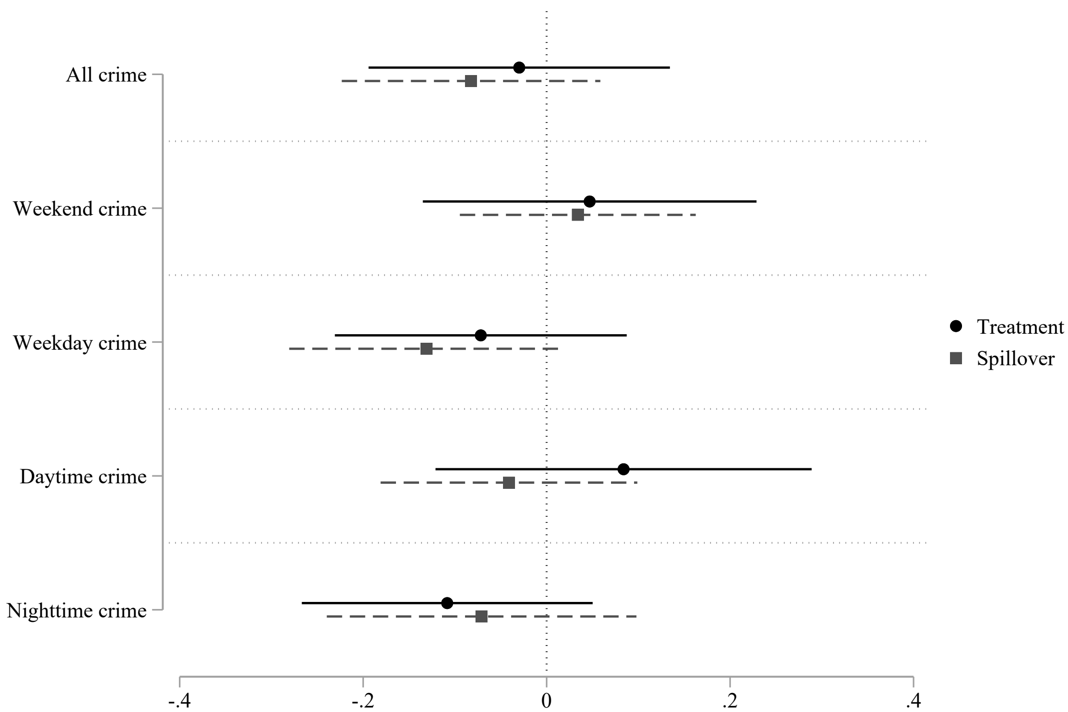
Notes: ITT on abuses by police (columns 1 and 3) and military (columns 2 and 4) based on monitoring (columns 1 and 2) and survey data (columns 3 and 4). Check marks indicate whether the corresponding coefficient remains statistically significant after applying the Benjamini-Hochberg or Holm-Bonferroni correction under the stringent assumption that each regression tests two hypotheses. 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 effects on crime using weighted standardized index**

	Admin data	
	Crime index	
	During intervention	After intervention
Treatment	-0.028	0.118
	(0.08)	(0.08)
Spillover	-0.080	0.073
	(0.07)	(0.07)
Individual controls	✗	✗
Neighborhood FE	✓	✓
Block-level controls	✓	✓
Observations	1167	1167
R^2	0.21	0.26
Control mean	0.089	-0.036

Notes: ITT on weighted standardized indices of crime during (column 1) and after (column 2) the intervention based on administrative data. Crimes are weighted by the average prison sentence under Colombian law. 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.

Figure A.6: **Treatment effects on crime by day and time using weighted standardized index**



Notes: ITT on weighted standardized indices of crime during the intervention based on administrative data, disaggregated by day and time. Crimes are weighted by the average prison sentence under Colombian law. All specifications include neighborhood fixed effects and block-level controls. Observations are weighted by the inverse probability of assignment to their realized treatment status. Lines denote 95% confidence intervals.

Table A.13: **Treatment effects on violent and non-violent crime in admin data**

	Admin data			
	During intervention		After intervention	
	Violent crime	Non-violent crime	Violent crime	Non-violent crime
Treatment	0.043*	-0.034	0.081*	0.032
	(0.02)	(0.03)	(0.05)	(0.03)
Spillover	0.004	-0.036	0.042	0.040*
	(0.02)	(0.02)	(0.04)	(0.02)
Individual controls	✗	✗	✗	✗
Neighborhood FE	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓
Observations	1167	1167	1167	1167
R^2	0.30	0.16	0.30	0.32
Control mean	0.080	0.080	0.102	0.058

Notes: ITT on violent and non-violent crime during (columns 1 and 2) and after (columns 3 and 4) the intervention based on administrative data. 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.14: **Treatment effects on violent crime victimization and witnessing in survey data**

	Survey data		
	Victim		**Witness**
	During intervention	After intervention	After intervention
Treatment	-0.016	-0.008	0.146***
	(0.05)	(0.05)	(0.05)
Spillover	-0.011	0.019	0.172***
	(0.04)	(0.04)	(0.04)
Individual-level controls	✓	✓	✓
Block-level controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Observations	7883	7883	7898
R^2	0.01	0.02	0.10
Control mean	-0.005	-0.009	-0.102

Notes: ITT on violent crime victimization during (columns 1 and 2) and after (columns 3 and 4) the intervention based on survey data and violent crime witnessing after the intervention (column 5) 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 are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.15: **Treatment effects on non-violent crime victimization and witnessing in survey data**

	Survey data		
	Victim		**Witness**
	During intervention	After intervention	After intervention
Treatment	0.014	-0.008	0.148***
	(0.04)	(0.04)	(0.05)
Spillover	0.041	0.006	0.180***
	(0.03)	(0.04)	(0.04)
Individual-level controls	✓	✓	✓
Block-level controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Observations	7865	7865	7840
R^2	0.03	0.03	0.11
Control mean	-0.024	-0.015	-0.118

Notes: ITT on non-violent crime victimization during (columns 1 and 2) and after (columns 3 and 4) the intervention based on survey data non-violent crime witnessing after the intervention (column 5) 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 are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.16: **Treatment effects on crime in admin data disaggregated by type**

	Admin data							
	During intervention				**After intervention**			
	Robbery	Homicide	Drug dealing	Possesion of firearms	Robbery	Homicide	Drug dealing	Possesion of firearms
Treatment	0.010	0.001	-0.009	0.003	0.098**	-0.002	0.004	0.010
	(0.03)	(0.01)	(0.01)	(0.01)	(0.05)	(0.01)	(0.00)	(0.01)
Spillover	-0.034	-0.004	-0.001	0.002	0.077*	-0.007	0.005	0.006
	(0.03)	(0.01)	(0.01)	(0.01)	(0.04)	(0.01)	(0.00)	(0.01)
Individual controls	✗	✗	✗	✗	✗	✗	✗	✗
Neighborhood FE	✓	✓	✓	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1167	1167	1167	1167	1167	1167	1167	1167
R^2	0.34	0.08	0.04	0.07	0.50	0.05	0.07	0.06
Control mean	0.131	0.011	0.015	0.004	0.131	0.018	0.004	0.007

Notes: ITT on crime during (columns 1 through 4) and after (columns 5 through 8) the intervention based on administrative data, disaggregated by type. 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.17: **Treatment effects on crime victimization in survey data disaggregated by type**

	Survey data								
	Crime victimization (during intervention)								
	Vandalism	Theft	Homicides	Attempted homicides	Armed robberie	Motor vehicle theft	Burglary	Gang activity	Extortion
Treatment	0.007	-0.003	0.003	-0.001	-0.004	-0.005	0.000	0.005**	0.000
	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Spillover	0.001	0.006	0.003**	-0.002	-0.002	-0.004	0.005***	0.004**	0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Individual controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	7896	7893	7903	7901	7895	7899	7902	7903	7906
R^2	0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.02	0.01
Control mean	0.013	0.021	0.001	0.005	0.019	0.013	0.004	0.002	0.001

Notes: ITT on crime victimization during the intervention, disaggregated by type. 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.17: **Treatment effects on crime victimization in survey data disaggregated by type (cont.)**

	Survey data								
	Crime victimization (after intervention)								
	Vandalism	Theft	Homicides	Attempted homicides	Armed robberie	Motor vehicle theft	Burglary	Gang activity	Extortion
Treatment	-0.014**	0.002	0.003	0.005	-0.009	-0.007	0.004	0.006	0.003
	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Spillover	-0.001	0.004	0.006*	0.004	-0.005	-0.003	0.001	-0.002	0.001
	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Individual controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	7896	7893	7903	7901	7895	7899	7902	7903	7906
R^2	0.02	0.03	0.01	0.01	0.02	0.01	0.01	0.01	0.01
Control mean	0.033	0.043	0.008	0.008	0.037	0.022	0.014	0.014	0.002

Notes: ITT on crime victimization after the intervention, disaggregated by type. 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.18: **Treatment effects on crime witnessing in survey data disaggregated by type**

	Survey data							
	Crime witnessing							
	Prostitution	Drug dealing	Drug consumption	Alcohol consumption	Vandalism	Theft	Homicide	Attempted homicide
Treatment	0.044	0.166***	0.142***	0.047	0.099**	0.077	0.158***	0.141***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Spillover	0.053	0.198***	0.210***	0.098**	0.129***	0.113***	0.170***	0.157***
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Individual controls	✓	✓	✓	✓	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓	✓	✓	✓
Observations	7903	7880	7899	7900	7903	7899	7901	7900
R^2	0.03	0.07	0.07	0.08	0.07	0.11	0.10	0.09
Control mean	-0.046	-0.127	-0.138	-0.061	-0.076	-0.081	-0.098	-0.083

Notes: ITT on crime witnessing after the intervention based on survey data, disaggregated by type. 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.18: **Treatment effects on crime witnessing in survey data disaggregated by type (cont.)**

	Survey data						
	Crime witnessing						
	Armed robbery	Motor vehicle theft	Burglary	Domestic violence	Gang activity	Possession of firearms	Extorsion
Treatment	0.073	0.097**	0.044	0.108**	0.151***	0.111**	0.095**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)
Spillover	0.110***	0.090**	0.071*	0.101***	0.148***	0.126***	0.095***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Individual controls	✓	✓	✓	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓	✓	✓
Observations	7902	7900	7900	7904	7887	7889	7891
R^2	0.11	0.06	0.07	0.06	0.10	0.12	0.04
Control mean	-0.077	-0.051	-0.059	-0.064	-0.095	-0.085	-0.056

Notes: ITT on crime witnessing after the intervention based on survey data, disaggregated by type. 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.

Figure A.7: **Alleged abuses by state security forces in Cali as reported to Attorney-General's Office, September 30, 2019–December 31, 2019**

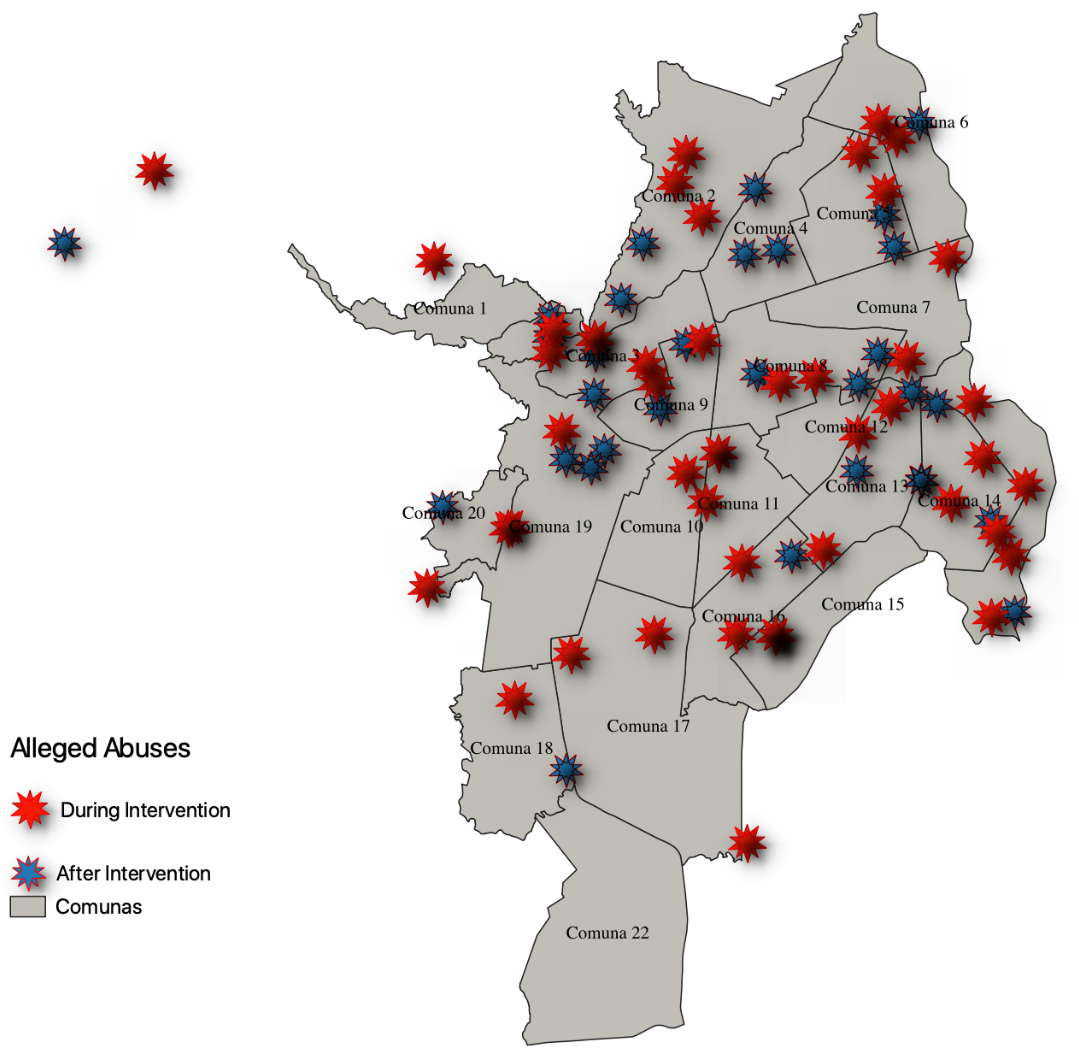


Figure A.8: **Alleged abuses by state security forces in *comuna* 18 as reported to Attorney-General's Office, September 30–November 18, 2019**

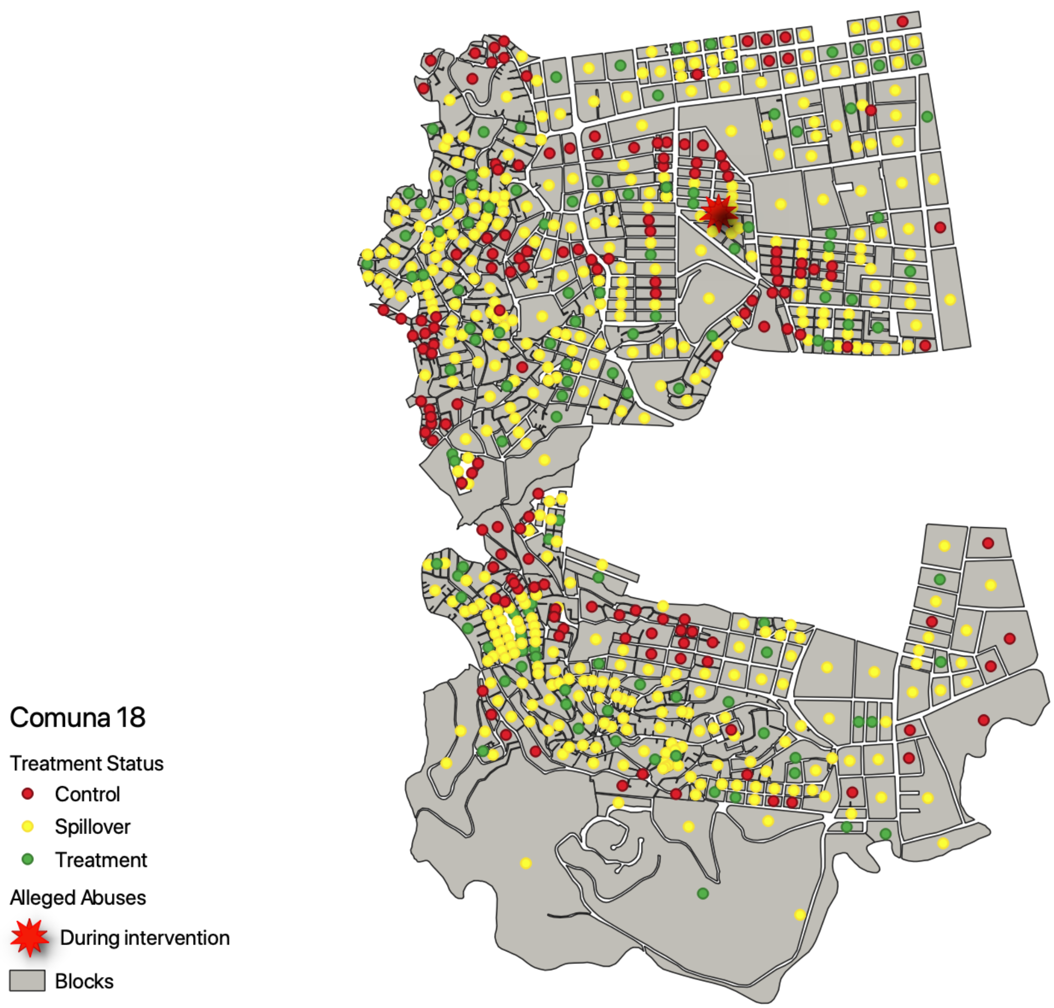


Figure A.9: **Alleged abuses by state security forces in *comuna* 20 as reported to Attorney General's Office, September 30–November 18, 2019)**

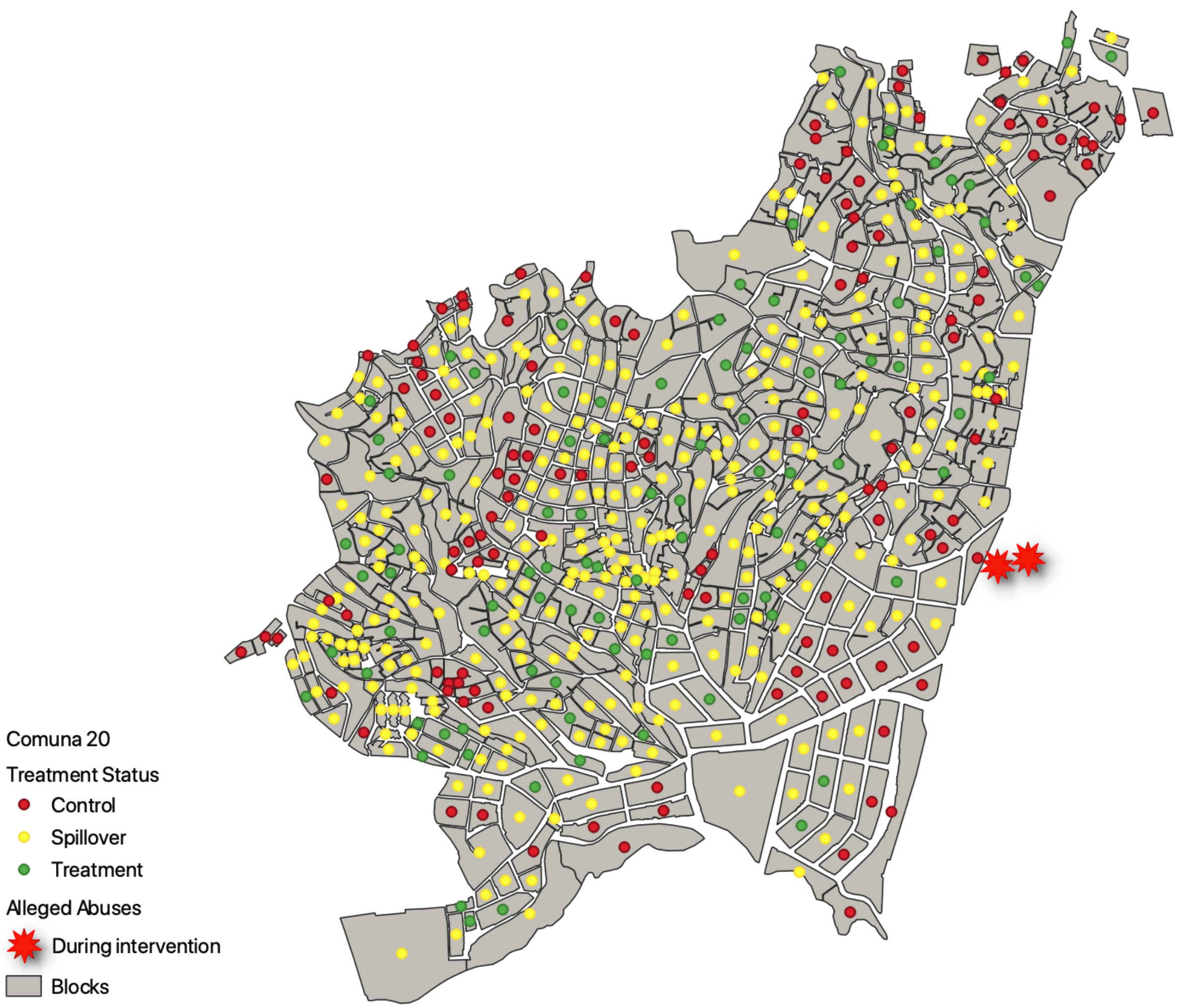


Figure A.10: **Military physical abuses in *comuna* 18 in monitoring survey**

Figure A.11: **Military physical abuses in *comuna* 20 in monitoring survey**

Figure A.12: **Military verbal abuses in *comuna* 18 in monitoring survey**

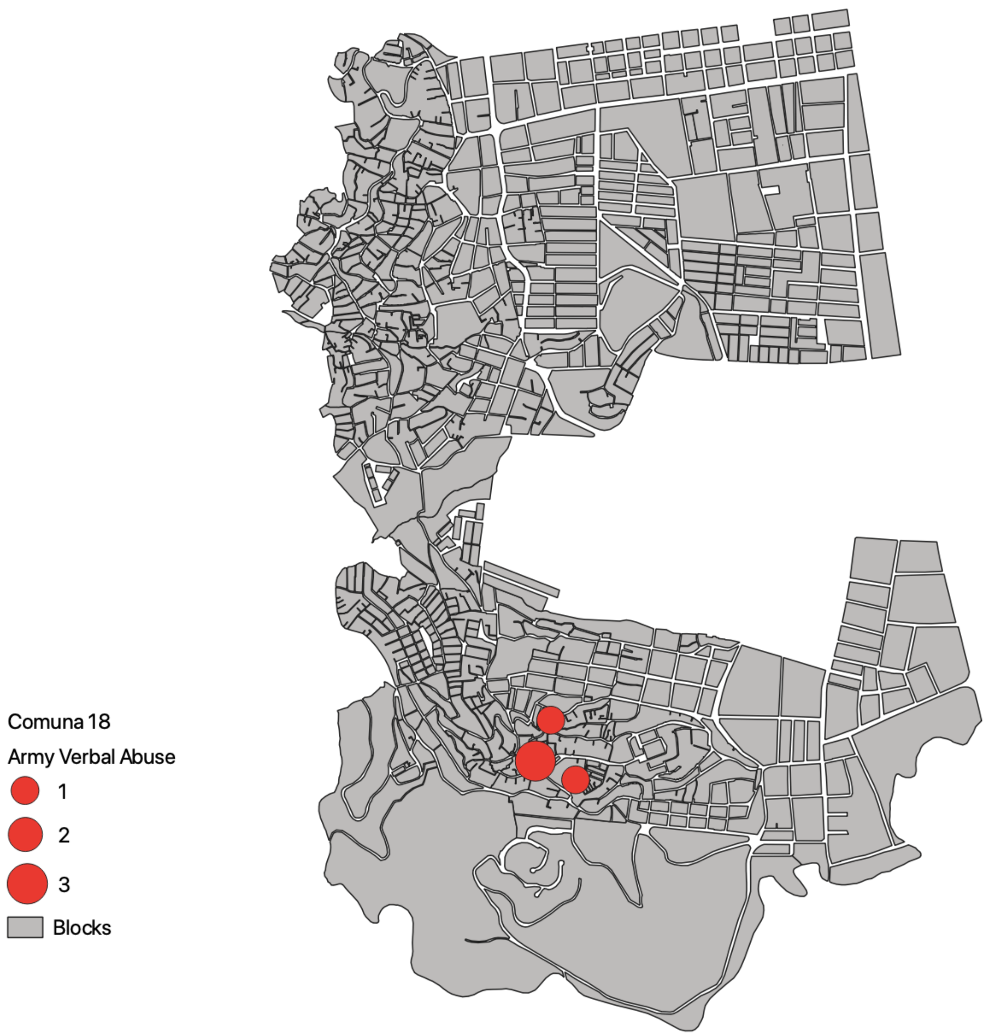


Figure A.13: **Military physical abuses in *comuna* 20 in monitoring survey**

Figure A.14: **Police physical abuses in *comuna* 18 in monitoring survey**

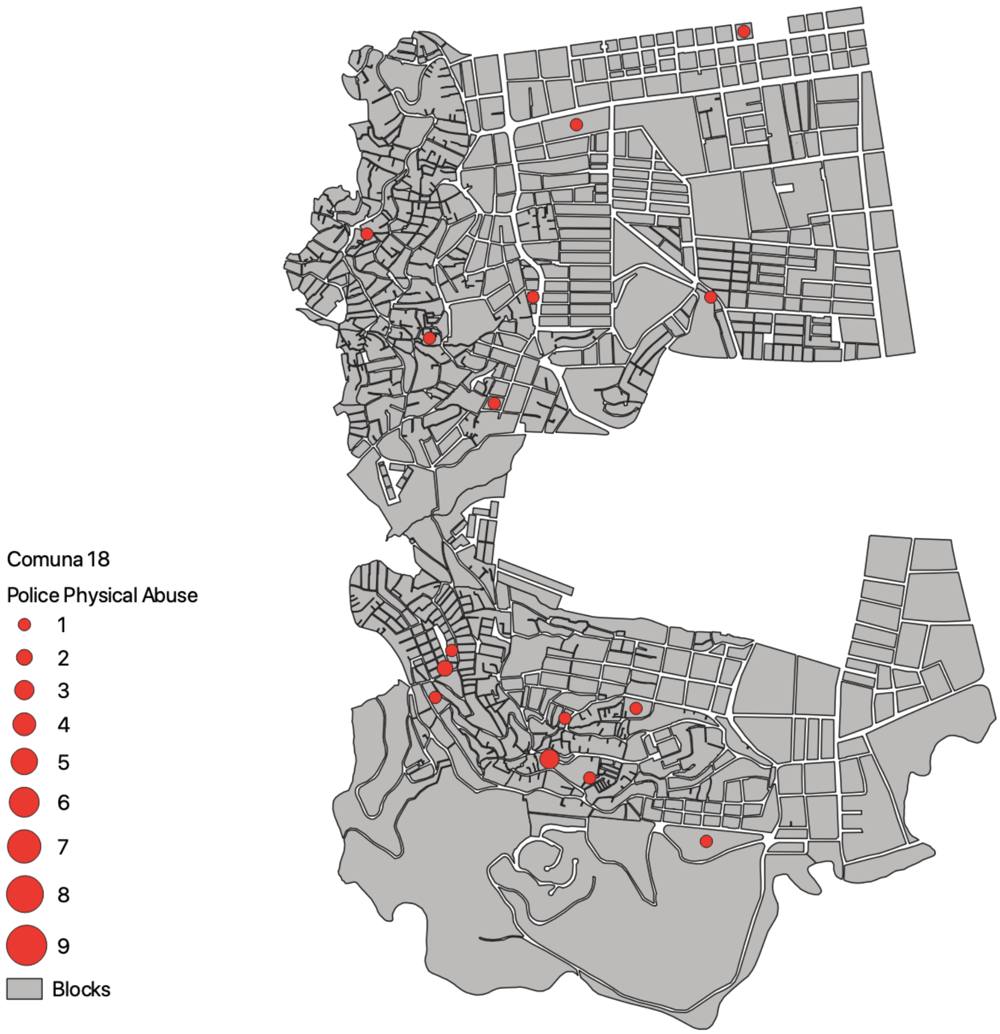


Figure A.15: **Police physical abuses in *comuna* 20 in monitoring survey**

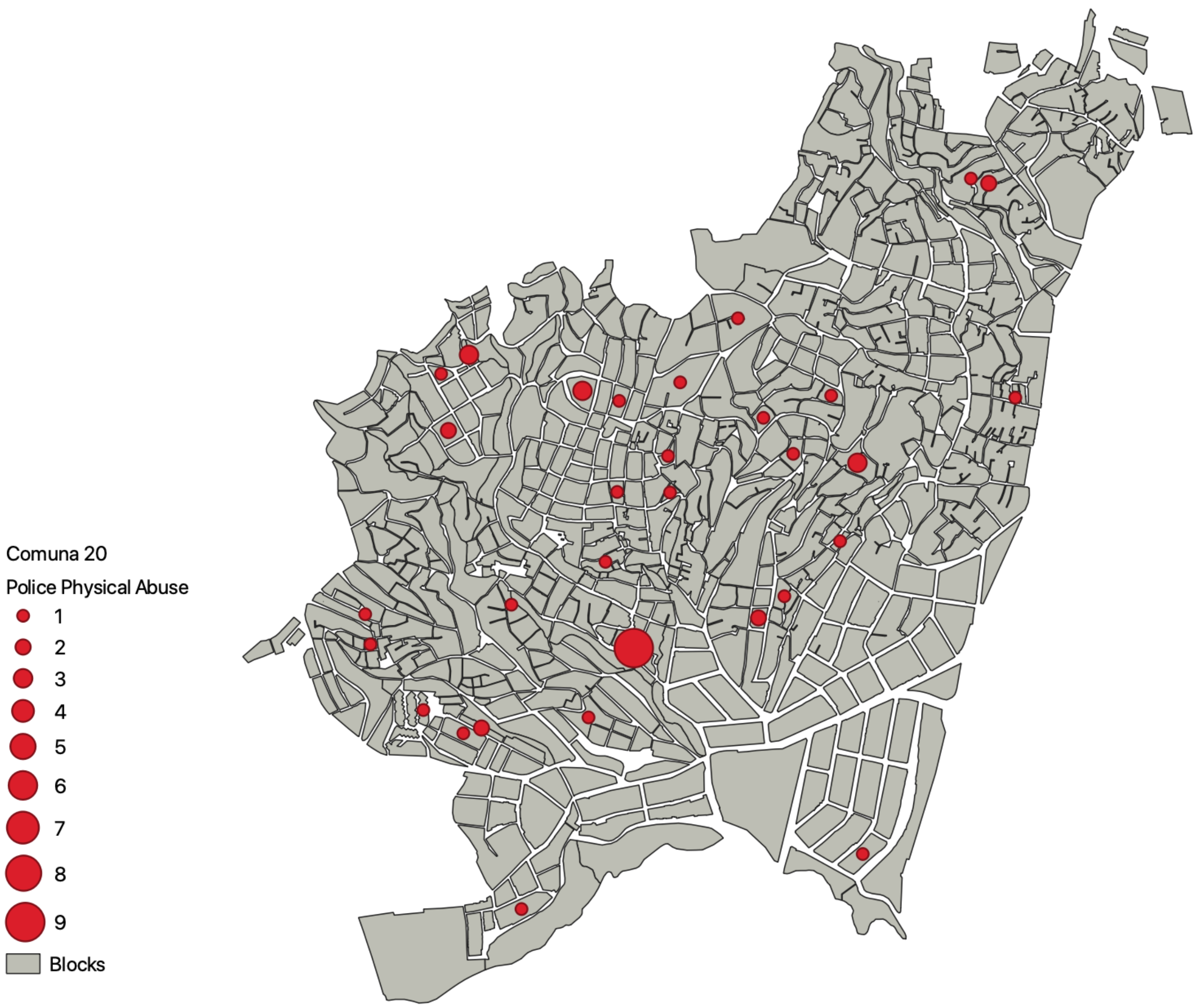


Figure A.16: **Police verbal abuses in *comuna* 18 in monitoring survey**

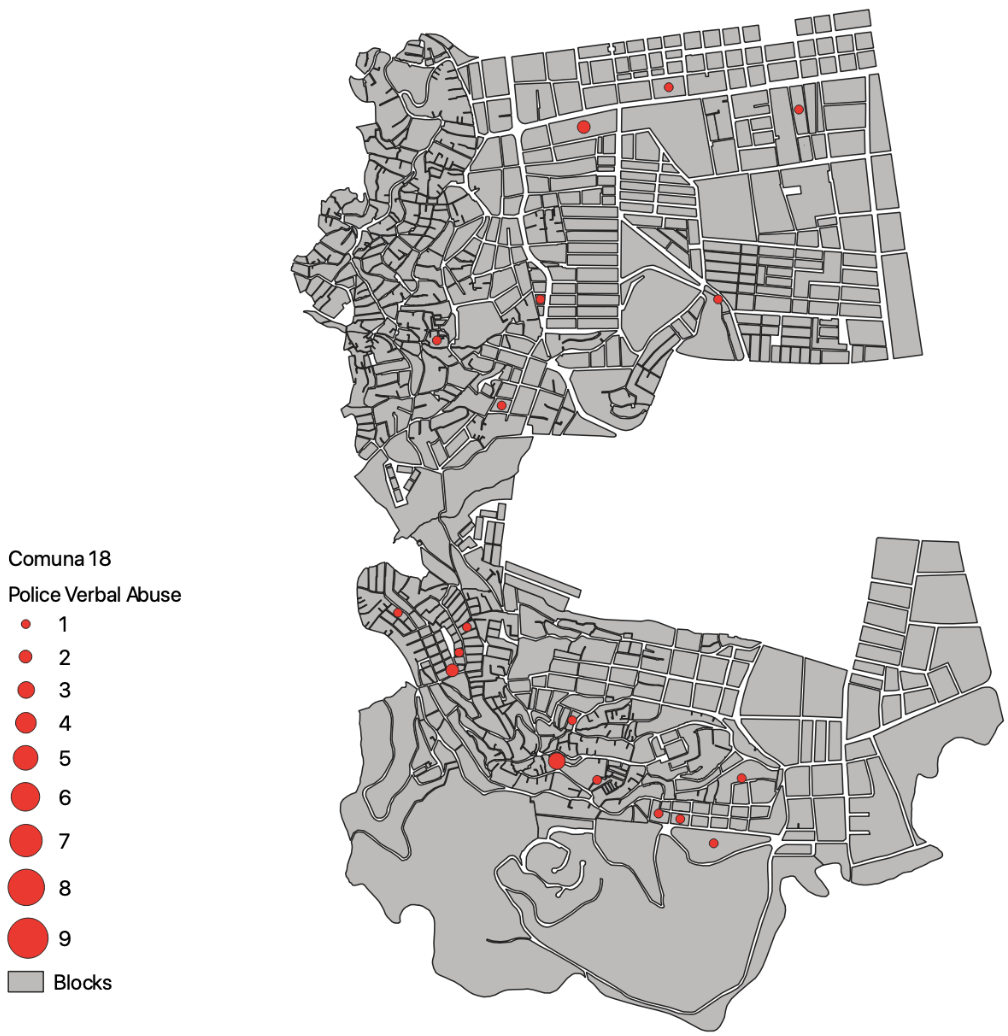


Figure A.17: **Police verbal abuses in *comuna* 20 in monitoring survey**

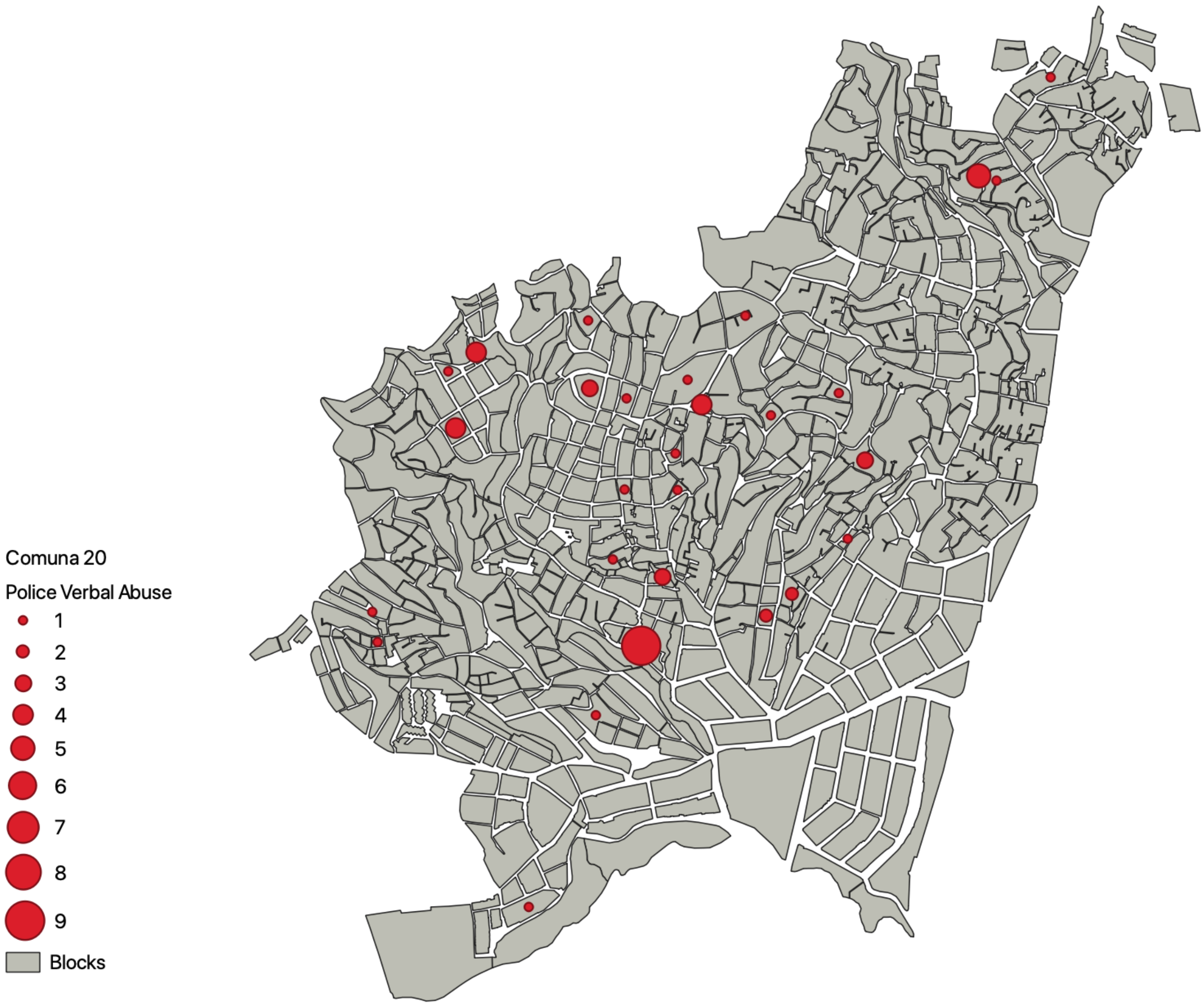


Table A.19: **Treatment effects on abuses disaggregated by date of survey**

	During intervention (before November 19)				**After intervention (after November 19)**			
	Physical abuse		**Verbal abuse**		**Physical abuse**		**Verbal abuse**	
	Police	Military	Police	Military	Police	Military	Police	Military
Treatment	0.041	0.000	0.055	0.000	0.039**	0.017*	0.032**	0.017*
	(0.04)	(.)	(0.05)	(.)	(0.02)	(0.01)	(0.01)	(0.01)
Spillover	-0.006	0.000	0.017	0.000	0.016	0.003	0.013	0.002
	(0.03)	(.)	(0.06)	(.)	(0.01)	(0.00)	(0.01)	(0.00)
Observations	850	850	850	850	1235	1235	1235	1235
R^2	0.04	.	0.06	.	0.04	0.05	0.04	0.04
Control mean	0.033	0.000	0.044	0.000	0.011	0.000	0.006	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. 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.20: **Treatment effects on abuses disaggregated by date of survey**

	During intervention with buffer (before December 2)				**After intervention with buffer (after December 2)**			
	Physical abuse		**Verbal abuse**		**Physical abuse**		**Verbal abuse**	
	Police	Military	Police	Military	Police	Military	Police	Military
Treatment	0.041	0.002	0.040	0.006	0.030*	0.015	0.037**	0.015
	(0.03)	(0.00)	(0.04)	(0.00)	(0.02)	(0.01)	(0.02)	(0.01)
Spillover	-0.008	-0.001	0.003	0.000	0.014	0.000	0.013	0.000
	(0.03)	(0.00)	(0.04)	(0.00)	(0.02)	(0.01)	(0.02)	(0.01)
Observations	1147	1147	1147	1147	938	938	938	938
R^2	0.04	0.02	0.05	0.03	0.04	0.05	0.04	0.05
Control mean	0.032	0.000	0.041	0.000	0.010	0.000	0.003	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. 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.21: **Heterogeneous treatment effects on crime, crime victimization, and crime witnessing by prior crime rate**

	Admin data		**Survey data**		
	Crime incidence		**Crime victimization**		**Crime witnessing**
	During intervention	After intervention	During intervention	After intervention	After intervention
Treatment	-0.003	0.019	0.002	0.003	0.184***
	(0.04)	(0.01)	(0.04)	(0.05)	(0.06)
Spillover	-0.023	0.018	0.040	0.048	0.218***
	(0.03)	(0.01)	(0.03)	(0.04)	(0.05)
Prior crime	0.097**	0.024	0.040	0.037	0.054**
	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)
Treatment × prior crime	-0.001	-0.024	-0.048	-0.038	-0.042
	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)
Spillover × prior crime	-0.034	-0.021	-0.035	-0.054	-0.045
	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)
Individual controls	✗	✗	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓
Observations	1167	1167	6449	7073	7837
R^2	0.33	0.28	0.03	0.03	0.12
Control mean	0.160	0.011	-0.008	-0.015	-0.119

Notes: HTE by prior crime rate on crime during (column 1) and after (column 2) the intervention based on administrative data; crime victimization during (column 3) and after (column 4) the intervention based on survey data; and crime witnessing after the intervention (column 5) based on survey data. All specifications include neighborhood fixed effects and block-level controls. Models 3-6 also include individual-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.22: **Heterogeneous treatment effects on crime victimization and crime witnessing by prior crime victimization**

	Survey data		
	Crime victimization		Crime witnessing
	During intervention	After intervention	During intervention
Treatment	0.004	-0.008	0.146***
	(0.04)	(0.05)	(0.05)
Spillover	0.025	0.013	0.182***
	(0.03)	(0.04)	(0.04)
Prior crime victimization	0.145***	0.057	0.177***
	(0.04)	(0.04)	(0.02)
Treatment × Prior crime victimization	0.042	0.078	0.046
	(0.08)	(0.07)	(0.04)
Spillover × Prior crime victimization	-0.005	-0.004	-0.052*
	(0.05)	(0.05)	(0.03)
Individual controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Block-level controls	✓	✓	✓
Observations	7845	7845	7800
R^2	0.05	0.04	0.15
Control mean	-0.021	-0.016	-0.119

Notes: HTE by prior crime victimization on crime victimization during (column 1) and after (column 2) the intervention based on survey data; and crime witnessing after the intervention (column 3) 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 are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.23: **Heterogeneous treatment effects on crime victimization and crime witnessing by gender**

	Survey data		
	Crime victimization		**Crime witnessing**
	During intervention	After intervention	During intervention
Treatment	-0.045	-0.068	0.144*
	(0.07)	(0.08)	(0.08)
Spillover	0.029	-0.042	0.134**
	(0.06)	(0.06)	(0.06)
Female	-0.042	-0.066	-0.041
	(0.06)	(0.06)	(0.05)
Treatment × Female	0.076	0.092	0.014
	(0.08)	(0.08)	(0.08)
Spillover × Female	-0.005	0.082	0.077
	(0.07)	(0.06)	(0.06)
Individual controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Block-level controls	✓	✓	✓
Observations	7845	7845	7837
R^2	0.03	0.03	0.12
Control mean	-0.021	-0.016	-0.119

Notes: HTE by gender on crime victimization during (column 1) and after (column 2) the intervention based on survey data; and crime witnessing after the intervention (column 3) 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 are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.24: **Heterogeneous treatment effects on perceptions of safety by prior crime rate**

	Survey data		
	All safety	**Personal safety**	**Business safety**
Treatment	-0.112**	-0.114**	0.217
	(0.05)	(0.05)	(0.14)
Spillover	-0.130***	-0.133***	0.017
	(0.04)	(0.04)	(0.13)
Prior crime	-0.083***	-0.083***	-0.065
	(0.02)	(0.02)	(0.05)
Treatment × prior crime	0.088***	0.087***	0.061
	(0.02)	(0.02)	(0.05)
Spillover × prior crime	0.092***	0.093***	0.068
	(0.03)	(0.03)	(0.06)
Individual-level controls	✓	✓	✓
Block-level controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Observations	7707	7708	1041
R^2	0.09	0.09	0.13
Control mean	0.077	0.078	-0.065

Notes: HTE by prior crime rate on perceptions of safety for all respondents (column 1), residents (column 2), and business owners (column 3) 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.

Table A.25: **Heterogeneous treatment effects on perceptions of safety by prior crime victimization**

	Survey data		
	All safety	**Personal safety**	**Business safety**
Treatment	-0.040	-0.042	0.308***
	(0.05)	(0.05)	(0.11)
Spillover	-0.063	-0.065*	0.083
	(0.04)	(0.04)	(0.10)
Prior crime victimization	-0.144***	-0.142***	-0.177***
	(0.03)	(0.03)	(0.07)
Treatment × Prior crime victimization	0.008	0.004	0.060
	(0.04)	(0.04)	(0.11)
Spillover × Prior crime victimization	0.027	0.025	0.096
	(0.04)	(0.03)	(0.08)
Individual-level controls	✓	✓	✓
Block-level controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Observations	7671	7672	1031
R^2	0.11	0.11	0.15
Control mean	0.077	0.078	-0.065

Notes: HTE by prior crime victimization on perceptions of safety for all respondents (column 1), residents (column 2), and business owners (column 3) 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.

Table A.26: **Heterogeneous treatment effects on perceptions of safety by gender**

	Survey data		
	All safety	**Personal safety**	**Business safety**
Treatment	-0.023	-0.027	0.340**
	(0.07)	(0.07)	(0.17)
Spillover	-0.004	-0.006	0.066
	(0.06)	(0.06)	(0.15)
Female	-0.166***	-0.166***	-0.078
	(0.06)	(0.06)	(0.18)
Treatment × female	-0.041	-0.038	-0.096
	(0.08)	(0.08)	(0.21)
Spillover × female	-0.092	-0.093	0.043
	(0.06)	(0.06)	(0.20)
Individual-level controls	✓	✓	✓
Block-level controls	✓	✓	✓
Neighborhood FE	✓	✓	✓
Observations	7707	7708	1041
R^2	0.09	0.09	0.13
Control mean	0.077	0.078	-0.065

Notes: HTE by gender on perceptions of safety for all respondents (column 1), residents (column 2), and business owners (column 3) 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.

Table A.27: **Heterogeneous treatment effects on abuses by prior crime rate**

	Monitoring data		Survey data	
	Police abuse	**Military abuse**	**Police abuse**	**Military abuse**
Treatment	0.032***	0.005	0.013	-0.001
	(0.01)	(0.00)	(0.01)	(0.00)
Spillover	0.001	0.001	0.033**	0.002
	(0.01)	(0.00)	(0.01)	(0.00)
Prior crime	-0.000	-0.000	-0.000	-0.000
	(0.00)	(0.00)	(0.01)	(0.00)
Treatment × prior crime	0.003	0.010	-0.003	-0.000
	(0.01)	(0.01)	(0.01)	(0.00)
Spillover × prior crime	0.024*	0.001	-0.005	-0.001
	(0.01)	(0.00)	(0.01)	(0.00)
Individual-level controls	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓
Observations	2085	2085	7908	7908
R^2	0.05	0.05	0.05	0.01
Control mean	0.015	0.000	0.114	0.012

Notes: HTE by prior crime rate 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.

Table A.28: **Heterogeneous treatment effects on abuses by prior crime victimization**

	Survey data	
	Police abuse	**Military abuse**
Treatment	0.009	-0.003
	(0.01)	(0.00)
Spillover	0.030**	0.002
	(0.01)	(0.00)
Prior crime victimization	0.025**	0.006
	(0.01)	(0.01)
Treatment × Prior crime victimization	0.005	0.004
	(0.02)	(0.01)
Spillover × Prior crime victimization	0.004	-0.002
	(0.01)	(0.01)
Individual-level controls	✓	✓
Block-level controls	✓	✓
Neighborhood FE	✓	✓
Observations	7870	7870
R^2	0.05	0.02
Control mean	0.114	0.012

Notes: HTE by prior crime victimization on abuses by police (column 1) and military (column 2) 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.

Table A.29: **Heterogeneous treatment effects on abuses by gender**

	Monitoring data		**Survey data**	
	Police abuse	**Military abuse**	**Police abuse**	**Military abuse**
Treatment	0.034**	0.008	0.009	0.002
	(0.02)	(0.01)	(0.02)	(0.01)
Spillover	0.011	0.006	0.035*	0.008*
	(0.02)	(0.01)	(0.02)	(0.00)
Female	0.002	0.000	-0.012	0.007
	(0.01)	(0.00)	(0.02)	(0.01)
Treatment × Female	-0.000	0.003	0.003	-0.006
	(0.02)	(0.01)	(0.03)	(0.01)
Spillover × Female	0.000	-0.007	-0.008	-0.009
	(0.02)	(0.01)	(0.02)	(0.01)
Individual-level controls	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓
Observations	2085	2085	7908	7908
R^2	0.05	0.04	0.05	0.01
Control mean	0.015	0.000	0.114	0.012

Notes: HTE by gender 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.

Table A.30: **Treatment effects on exposure to police and military**

	Monitoring data		**Survey data**	
	During intervention		**After intervention**	
	Seen on block		**Seen on block**	
	Police	Military	Police	Military
Treatment	0.025	0.098***	0.044**	0.023
	(0.04)	(0.03)	(0.02)	(0.02)
Spillover	0.027	0.075*	0.034**	0.005
	(0.05)	(0.04)	(0.02)	(0.02)
Individual-level controls	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓
Observations	2085	2085	7908	7908
R^2	0.06	0.10	0.04	0.13
Control Mean	0.612	0.257	0.788	0.387

Notes: ITT on exposure to police and military during (columns 1 and 2) and after (columns 3 and 4) the intervention 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.

Table A.31: **Treatment effects on police and military activities**

	Monitoring data						**Survey data**					
	During intervention						**After intervention**					
	Seen making arrests		**Seen checking ID**		**Seen talking with citizens**		**Seen making arrests**		**Seen checking ID**		**Seen talking with citizens**	
	Police	Military	Police	Military	Police	Military	Police	Military	Police	Military	Police	Military
Treatment	0.032**	0.010*	0.030	0.009	0.004	0.012	0.059***	0.011	0.035	0.007	-0.026	0.006
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Spillover	0.029*	0.004	0.037	0.011	0.021	-0.004	0.062***	0.019**	0.045***	0.021*	0.004	0.009
	(0.02)	(0.00)	(0.03)	(0.01)	(0.03)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
Individual-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Block-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Neighborhood FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	2085	2085	2085	2085	2085	2085	7908	7908	7908	7908	7908	7908
R^2	0.05	0.04	0.05	0.05	0.03	0.05	0.05	0.02	0.04	0.05	0.03	0.08
Control Mean	0.034	0.000	0.118	0.039	0.128	0.045	0.302	0.065	0.386	0.129	0.521	0.193

Notes: ITT on police and military activities during (columns 1-6) and after (columns 7-12) the intervention 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.

Table A.32: **Treatment effects on cooperation with police and military**

	Survey data		
	Cooperation with authorities		
	All	Police	Military
Treatment	0.086**	0.080*	0.069*
	(0.04)	(0.04)	(0.04)
Spillover	0.087**	0.091**	0.034
	(0.03)	(0.04)	(0.03)
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 with the police (column 2) and military (column 3) separately 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.