# Little evidence that military policing reduces crime or improves human security

Robert A. Blair[1]
Michael Weintraub[2]

*Nature Human Behaviour*, forthcoming

[1]Corresponding author. Associate Professor, Department of Political Science and Watson Institute for International and Public Affairs, Brown University, Providence, RI, USA, robert_blair@brown.edu
[2]Corresponding author. Associate Professor, School of Government, Universidad de los Andes, Bogotá, Colombia, mlw@uniandes.edu.co

**Abstract**

Governments in low and middle income countries routinely deploy their armed forces for domestic policing operations. Advocates of these policies claim they reduce crime, while detractors argue they undermine human rights. We experimentally evaluate a military policing intervention in Cali, Colombia. 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 residents, and firsthand observations from civilian monitors, we find little to no credible evidence that military policing reduced crime or improved perceptions of safety during the intervention. If anything, we find that military policing likely exacerbated crime after the intervention was complete. We also find evidence of increased human rights abuses in our survey data (though not in the administrative data or in the firsthand observations of civilian monitors), largely committed by police officers rather than soldiers. We argue the benefits of military policing are likely small and not worth the costs.

# 1 INTRODUCTION

Governments in low and middle income countries 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 [1, 2]. Military policing is increasingly common outside of Latin America as well, for example in Indonesia [3], the Philippines [4], and South Africa [5]. Even in the US, commentators occasionally urge the deployment of troops to support police departments in "high-crime, drug-infested urban areas" [6, p. 220].

In light of the widespread use of military forces in domestic policing operations and ongoing arguments about the benefits of this approach, we examine the effectiveness and broader implications of military policing through an experimental evaluation of the Plan Fortaleza program in Cali, Colombia. The program involved recurring, intensive military patrols targeting hot spots for crime. Military policing programs like Plan Fortaleza draw on some of the same theories that motivate other place-based policing strategies [7–11], but adapt them to contexts where the assumptions underlying those theories break down.

Place-based policing hinges crucially on the assumption that police presence deters criminals [12, 13]. Proponents of military policing implicitly or explicitly call this assumption into question. In many Latin American countries, the police are poorly trained and equipped and have a reputation for corruption and complicity with organized crime [14–16]. If deterrence depends on the certainty, severity, and swiftness of apprehension and incarceration [17–21], then corrupt, collusive, and resource constrained police forces are unlikely to deter and may in fact embolden criminals.

Supporters of military policing argue that because soldiers are better trained and have greater logistical and coercive capacity than police officers, they should be more effective at preventing crime. Soldiers are also typically subject to more stringent accountability mechanisms [22], which may reduce the risk that misconduct goes undetected and unpunished. Moreover, in many low and middle income countries, and in Latin America in particular, citizens perceive the military as more trustworthy and respectful of human rights than the police [16, 23]. Advocates cite these differences to argue that soldiers should be more effective than police officers at instilling perceptions of safety while simultaneously protecting citizens from abuse (see [24] for a summary of this debate).

But these arguments are hotly contested. Opponents of military policing counter that deterrence depends on the ability to investigate crimes [25], which soldiers are not trained to do. Organized criminal groups may respond violently to the military's presence [26], thus exacerbating crime and eroding perceptions of security [24, 27]. Most important, critics warn that soldiers are socialized to use force in ways that police officers are not, potentially undermining human rights [6, 28]. The risk of human rights abuses may be especially high when soldiers are stationed in densely populated urban areas [15].

To date, however, arguments on both sides of this debate remain almost entirely anecdotal and impressionistic. Despite the increasing prevalence of military policing in low and middle income countries, empirical evidence of its efficacy remains scarce. The expansive literature on hot spots policing is based almost exclusively in high income countries, and generally does not address militarization [29, 30]. Most studies of militarization focus on SWAT team deployments or transfers of military hardware to police departments in the US, with mixed results [31–36]. Other studies explore militarization of the police in Latin America, but do not address policing by the military per se [37–39].

This is a crucial distinction, since militaries tend to differ dramatically from even the most heavily militarized police forces, in particular in the expectation that they will use force not to "serve and protect" civilians but to "overwhelm and defeat" enemies on the battlefield [28, p. 329]. Only a small handful of studies have tested the effects of "constabularizing" the military for purposes of law enforcement, all using observational data [2, 24, 27]. As informative as these studies have been, observational research on military policing must overcome enormous inferential challenges. Quasi-experimental evidence on the efficacy of military policing is rare, and experimental evidence is, to our knowledge, non-existent.

Our experimental evaluation focuses on Cali, Colombia, the country's third largest city and one of its most violent. In 2018, the year before our study began, Cali recorded 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 deployed nightly military patrols to two "comunas" (communes) with some of the highest homicide rates in the city. We partnered with the Mayor's Office, the Third Brigade of the Colombian Armed Forces, and Innovations for Poverty Action (IPA) Colombia to evaluate the impact of these operations using a randomized controlled trial, with treatment assigned at the level of the "manzana" (city block).

Our sample consists of 1,255 blocks, 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 Supplementary Information E.) To evaluate the program, we combine timestamped, geolocated administrative data on crime and human rights abuses with two waves of surveys reaching over 10,000 respondents in total. We complement these data with detailed firsthand observations from civilian monitors hired to accompany the soldiers while on patrol. Our study was approved by the Ethics Committee at Universidad de los Andes (Acta 1073 de 2019 and Acta 1034 de 2019). We discuss the ethics of the Plan Fortaleza program and our evaluation of it in Section 4.1.

Our results suggest that military policing in Cali was at best ineffective and at worst counterproductive. We find little to no evidence that Plan Fortaleza reduced crime in the administrative data while the intervention was ongoin, and if anything our results suggest that it exacerbated crime after the intervention was complete. We observe an increase in crime in the administrative data after the intervention alongside an increase in citizens' accounts of witnessing and reporting crimes and an increase in the frequency of arrests. These seemingly adverse effects on crime do not appear to be artifacts of heightened vigilance on the part of civilians.

Perhaps relatedly, we find little to no evidence that military policing improved perceptions of safety, except perhaps among business owners. We also find some evidence of increased human rights abuses, but only in our surveys. In most 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 criminals.) But this result is somewhat 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. Nonetheless, taken together, our results suggest that governments should seek other strategies for curbing crime in the world's most violent cities.

As with any study focused on a single case, we cannot be sure how far our results will generalize. Cali does, however, share a number of traits with other cities in Latin America, and Plan Fortaleza resembles military policing programs in other contexts. As in many Latin American cities, access to security and other services in Cali depends on where one lives and the socioeconomic class to which one belongs. The police in Cali have a troubled history involving corruption, collusion, 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 in the US and elsewhere, including Colombia [1, 40]. Almost all studies of hot spots policing focus on particular neighborhoods within particular cities [30]. Our study of military policing uses a similar design.

Cali is also an important test case for military policing because it remains one of the most dangerous cities in the world, as we show in Figure 1, with gangs competing for local extortion, drug trafficking, and money laundering opportunities. 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 authorities. Given the city's ongoing struggle to control 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 other cities in Latin America and beyond, where military policing is routinely used but seldom rigorously evaluated.

# 2 Results

## 2.1 Crime, crime victimization, and crime witnessing

Advocates view military policing as a necessary temporary measure to deter crime. Table 1 tests this proposition by reporting the intention-to-treat (ITT) effect of the Plan Fortaleza program on an index of crimes committed during (column 1) and after (column 2) the intervention using administrative data. The index comprises murders, armed robberies, thefts, illegal drug sales, and illegal possession of weapons. We aggregate these crimes into an additive index at the block level. We report randomization inference (RI) $p$-values alongside more conventional $p$-values for all analyses; the former are generally more conservative than the latter. For ITT estimates with $p$-values greater than 0.1 we also report Bayes factors ($BF_{10}$) to help us identify possible Type II errors. We document deviations from our pre-analysis plan (PAP, available at https://osf.io/95cz3?mode=&revisionId=&view_only=) in Section 4.8.

Contrary to the arguments of advocates, we find little to no credible evidence that Plan Fortaleza reduced the prevalence of crime on treatment blocks while the intervention was ongoing ($\beta = 0.003$, CI $= [-0.068, 0.074]$, $p = 0.934$, RI $p = 0.959$, $BF_{10} = 0.107$). This null effect is unlikely to be an artifact of insufficient statistical power: as we discuss in Supplementary Information C, we are powered to detect even small changes in crime in the administrative data, and even smaller changes in crime victimization and witnessing in the survey. The null is also unlikely to be a Type II error: the corresponding Bayes factor is well below the $1/3$ threshold typically used to indicate support for the null hypothesis [41–43]. Nor do we find any credible evidence that Plan

Fortaleza reduced crime on spillover blocks while the intervention was ongoing ($\beta = -0.038$, CI $= [-0.097, 0.022]$, $p = 0.212$, RI $p = 0.411$, $BF_{10} = 0.577$).

Also contrary to proponents' claims, if anything we find that Plan Fortaleza exacerbated crime after the intervention was complete. Relative to control blocks, we observe 0.110 more crimes on treatment blocks between the end of the intervention on November 19, 2019 and the end of the year ($\beta = 0.110$, CI $= [0.011, 0.208]$, $p = 0.029$, RI $p = 0.136$). While this effect is no longer statistically significant when using RI $p$-values, it constitutes a substantively large increase of 69% relative to the control group mean (0.160 crimes per block after the intervention was complete), implying 24 more crimes in the treatment group distributed across 214 blocks. We also observe 0.083 more crimes on spillover blocks after the intervention was complete, though this ITT is not quite statistically significant at conventional levels ($\beta = 0.083$, CI $= [-0.003, 0.169]$, $p = 0.059$, RI $p = 0.138$).

Table 1 also reports the ITT on crime victimization in the endline survey (columns 3 and 4). Respondents were asked if they or someone in their household had been victimized by any of 10 crimes in the past six months: vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, domestic violence, or extortion. Respondents were also asked the month in which each crime occurred, and, for crimes committed in November 2019, whether they occurred before or after the massive nationwide protests that coincided with the end of the intervention (as described in Section 4 below). We code indicators for each crime, then aggregate them into a standardized additive index.

We find little to no credible evidence that Plan Fortaleza reduced crime victimization on treatment ($\beta = 0.006$, CI $= [-0.077, 0.089]$, $p = 0.886$, RI $p = 0.927$, $BF_{10} = 0.045$) or spillover blocks ($\beta = 0.026$, CI $= [-0.034, 0.086]$, $p = 0.389$, RI $p = 0.610$, $BF_{10} = 0.055$) during the intervention, or on treatment ($\beta = -0.007$, CI $= [-0.098, 0.085]$, $p = 0.886$, RI $p = 0.914$, $BF_{10} = 0.042$) or spillover blocks ($\beta = 0.013$, CI $= [-0.061, 0.087]$, $p = 0.729$, RI $p = 0.802$, $BF_{10} = 0.051$) after the intervention was complete, with Bayes factors indicative of evidence in favor of the null. While the null effects on crime victimization after the intervention are inconsistent with the apparent increase in crime in the administrative data, this discrepancy may reflect the fact that randomly selected survey respondents were relatively unlikely to have been victims of crime in the recent past. We consider this and other possible interpretations in the discussion.

Finally, Table 1 reports the ITT on residents' reports of witnessing crime in the endline survey (column 5). In addition to the 10 crimes listed above, respondents were asked how frequently they had witnessed prostitution, illegal drug sales, illegal drug consumption, public alcohol consumption, or illegal possession of firearms in the previous month. Frequency was measured on a Likert scale from 1 to 4; we aggregate these reports into a standardized additive index. We find that Plan Fortaleza increased reports of witnessing crimes by 0.153 standard deviations on treatment blocks ($\beta = 0.153$, CI $= [0.051, 0.256]$, $p = 0.003$, RI $p = 0.038$) and by 0.186 standard deviations on spillover blocks ($\beta = 0.186$, CI $= [0.101, 0.270]$, $p < 0.001$, RI $p = 0.001$) after the intervention.

As we show in Supplementary Information G.3, these ITTs hold across almost all categories of crime witnessing in the index, though levels of statistical significance vary. In exploratory analyses in Supplementary Information J.2 and J.3 we further show that endline survey respondents were more likely to report crimes and otherwise cooperate with the authorities on spillover blocks ($\beta = 0.091$, CI $= [0.021, 0.161]$, $p = 0.011$, RI $p = 0.061$); respondents were also more likely to report crimes on treatment blocks, though this ITT is not quite statistically significant at conventional levels ($\beta = 0.080$, CI $= [-0.005, 0.165]$, $p = 0.066$, RI $p = 0.148$). We also show that endline survey respondents were more likely to observe police officers making arrests on both treatment ($\beta = 0.059$, CI $= [0.015, 0.103]$, $p = 0.008$, RI $p = 0.068$) and spillover blocks ($\beta = 0.062$, CI $= [0.027, 0.097]$, $p < 0.001$, RI $p = 0.013$) after the intervention. These results are all consistent with our finding in Table 1 that crime increased after the intervention was complete.

## 2.2 CRIME DISAGGREGATED BY DAY AND TIME

All Plan Fortaleza patrols occurred on weekday nights. Following our PAP, in Figure 2 we use administrative data to distinguish crimes committed when soldiers were and were not physically present on the streets. We find little to no credible evidence of reduced crime on treatment blocks while the intervention was ongoing, even on weekdays ($\beta = 0.006$, CI = $[-0.053, 0.066]$, $p = 0.834$, RI $p = 0.901$, $BF_{10} = 0.104$) and at night ($\beta = -0.008$, CI = $[-0.054, 0.038]$, $p = 0.741$, RI $p = 0.812$, $BF_{10} = 0.123$), when soldiers were on patrol. We similarly find little to no credible evidence of reduced crime on spillover blocks on weekdays ($\beta = -0.038$, CI = $[-0.088, 0.011]$, $p = 0.129$, RI $p = 0.365$, $BF_{10} = 0.907$) or at night ($\beta = -0.017$, CI = $[-0.060, 0.025]$, $p = 0.424$, RI $p = 0.514$, $BF_{10} = 0.175$).

## 2.3 PERCEPTIONS OF SAFETY

Skeptics of place-based policing worry that citizens will interpret increased police presence as a signal that their neighborhoods are unsafe [44, 45]. While existing studies seem to belie this concern [46, 47], they focus exclusively on high income countries, and on place-based strategies implemented by police officers rather than soldiers. The effects of military policing on perceptions of safety are, to our knowledge, unknown. Figure 3 reports the ITT of the Plan Fortaleza program on perceptions of safety in our endline survey. We report results for all respondents together (row 1) and for residents and business owners separately (rows 2 and 3, respectively). This last analysis should be interpreted as exploratory, as we did not preregister hypotheses regarding business owners specifically.

Respondents were asked how safe they feel talking on a smartphone or walking their blocks during the day and at night, and how worried they are about becoming victims of violent or non-violent crime in the next two weeks. Respondents were also asked about precautions they had taken for fear of crime in the previous month, including avoiding public transportation, staying home at night, changing schools or jobs, 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.

We find little to no credible evidence that Plan Fortaleza improved perceptions of safety among residents: the ITT is negative and not statistically significant on either treatment ($\beta = -0.052$, CI $= [-0.144, 0.041]$, $p = 0.271$, RI $p = 0.431$, $BF_{10} = 0.237$) or spillover blocks ($\beta = -0.068$, CI $= [-0.146, 0.010]$, $p = 0.089$, RI $p = 0.186$). We do, however, find some evidence that the program improved perceptions of safety among business owners; this effect is substantively large, but only weakly statistically significant when using RI $p$-values ($\beta = 0.284$, CI $= [0.056, 0.512]$, $p = 0.015$, RI $p = 0.064$). One possible explanation for this discrepancy between business owners and randomly selected residents is that the former were more likely than the latter to be physically present during patrols, and more likely to interact with soldiers, who sometimes purchased food or water from local businesses. Another possible explanation is that business owners were more sensitive to threats posed by gangs (e.g. extortion), and thus more receptive to military patrols. These explanations are speculative, however, and overall the effects on perceived safety are mixed at best.

### 2.4 ABUSES

Opponents of military policing express particular concern about its potential adverse effects on human rights abuses [2, 6, 24, 28]. Figure 4 reports the ITT of the Plan Fortaleza program on abuses reported by monitoring (top panel) and endline survey respondents (bottom panel). 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 by police officers or soldiers in the past month. Since most respondents who reported physical abuse also reported verbal abuse, we collapse the two categories into a single indicator. While the monitoring survey continued for roughly a month after the intervention was complete, to allow for telescoping we assume that any abuses reported by monitoring survey respondents occurred while the intervention was ongoing. We assume that any abuses reported by endline survey respondents occurred after the intervention was complete. In Supplementary Information H we probe the robustness of our results to alternative coding rules.

We find little to no credible evidence that Plan Fortaleza increased the prevalence of abuses by soldiers during the intervention according to the monitoring survey: while the ITTs are positive on both treatment ($\beta = 0.010$, CI = $[-0.001, 0.022]$, $p = 0.071$, RI $p = 0.430$) and spillover blocks ($\beta = 0.001$, CI = $[-0.005, 0.007]$, $p = 0.778$, RI $p = 0.942$, $B_{10} = 0.110$), they are substantively small and not statistically significant at conventional levels. Importantly, reports of military abuse were exceedingly rare: just 10 out of 2,085 monitoring survey respondents (0.48% of the sample) reported either verbal or physical abuse by soldiers. Equally importantly, none of these 10 respondents was surveyed while implementation was ongoing. (They were surveyed on or after November 29, 11 days after the end of the intervention.) While we allow for telescoping, it is unclear if these reports were related to Plan Fortaleza. We similarly find little to no credible evidence of increased military abuse in the endline survey after the intervention was complete, either on treatment ($\beta = -0.001$, CI = $[-0.009, 0.006]$, $p = 0.716$, RI $p = 0.783$, $B_{10} = 0.075$) or spillover blocks ($\beta = 0.002$, CI = $[-0.005, 0.009]$, $p = 0.544$, RI $p = 0.613$, $B_{10} = 0.033$), with Bayes factors indicative of evidence in favor of the null.

We find more robust evidence of increased abuses by the police. Compared to control blocks, monitoring survey respondents on treatment blocks were 0.037 percentage points more likely to report abuse by police officers during the intervention ($\beta = 0.037$, CI = $[0.016, 0.058]$, $p = 0.001$, RI $p = 0.103$), though this effect is no longer statistically significant at conventional levels when using RI $p$-values. We find little to no credible evidence of increased police abuse on spillover blocks during the intervention ($\beta = 0.016$, CI = $[-0.005, 0.037]$, $p = 0.137$, RI $p = 0.385$, $BF_{10} = 0.229$). Police abuse was more common than military abuse overall, with 72 monitoring survey respondents (3.45% of the sample) reporting at least one incident of physical or verbal abuse by a police officer. Roughly half of monitoring survey respondents who reported police abuse were surveyed while the intervention was ongoing, and roughly half were surveyed after.

We also find some evidence that increased police abuse persisted over time, as residents of spillover blocks were 0.030 percentage points more likely to report abuses by police officers in the endline survey ($\beta = 0.030$, CI = $[0.007, 0.053]$, $p = 0.011$, RI $p = 0.028$). Residents of treatment blocks were also more likely to report police abuse at endline, though the ITT is substantively small and not statistically significant, and the Bayes factor is indicative of evidence in favor of the null ($\beta = 0.011$, CI = $[-0.015, 0.037]$, $p = 0.417$, RI $p = 0.541$, $B_{10} = 0.041$). It is not clear why police abuse would have shifted from treatment to spillover blocks over time, though the high density of these neighborhoods and the close proximity of treatment to spillover blocks suggests that a shift of this sort is not altogether surprising.

With just one exception, all monitoring survey respondents who reported military abuse also reported police abuse. This raises the possibility that civilians mistook police officers for soldiers, and so misreported the perpetrators of the abuses they witnessed. But this strikes us as unlikely. Colombian police officers do not at all resemble the soldiers that participated in Plan Fortaleza. This is true even of the more heavily militarized units of the Colombian police, such as the Mobile Anti-Disturbance Squadron (known by its Spanish acronym, ESMAD). While on patrol, soldiers wore bullet proof vests and traveled in vehicles that were clearly and conspicuously marked with the word "ejército" (army). Moreover, as we show in Supplementary Information J.1, monitoring survey respondents were more likely to report military presence on treatment blocks during the intervention ($\beta = 0.105$, CI = $[0.034, 0.175]$, $p = 0.004$, RI $p = 0.064$) but no more likely to report police presence ($\beta = 0.032$, CI = $[-0.049, 0.112]$, $p = 0.439$, RI $p = 0.611$, $B_{10} = 0.067$). If residents mistook soldiers for police officers, then intuitively they should have reported an increase in both military and police presence. But they did not.

In contrast to our monitoring survey results, we find little to no credible 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. The Attorney-General's data similarly includes only one allegation that occurred within our two study communes during the period of the intervention. The incident involved a transit police officer who was accused of unfairly restricting a citizen's freedom of movement, probably in relation to Cali's existing traffic laws. It seems unlikely that this incident was related to Plan Fortaleza in any direct way. Another two allegations of abuse occurred after the intervention was complete, both involving police officers. Given the nature and timing of these incidents—which we describe in detail in Supplementary Information H—it seems similarly unlikely that they were related to Plan Fortaleza.

## 3 Discussion

What explains the null or even adverse effects of Cali's military policing program, especially after the intervention was complete? There are several possible explanations. One is that police officers abandoned treatment and spillover blocks, perhaps because they assumed (incorrectly) that soldiers would permanently replace them. Our results, however, are not consistent with this explanation. As we show in Supplementary Information J.1 and J.2, if anything we observe greater police presence on both treatment and spillover blocks after the intervention was complete. We also observe more frequent arrests by police officers on both treatment and spillover blocks after the intervention. 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 unsafe), and that the crimes they witnessed and reported were subsequently entered into the city's administrative records. 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 apparently adverse effect on crime is illusory, masking a more benign—even beneficial—increase in cooperation between civilians and the police.

While this explanation strikes us as plausible, it is inconsistent with some of our 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 residents witnessed and the types of crimes that appeared in the city's administrative records. But we do not. As we show in Supplementary Information G.3, there is very little correlation between the types of crimes that appeared in the administrative data (e.g. theft and armed robbery) and the types of crimes that respondents reported witnessing in the endline survey (e.g. homicides, drug dealing, and illegal possession of firearms). Heightened vigilance might explain the positive effect on witnessing these latter crimes in the survey, but it cannot explain the null effect on the prevalence of these same crimes in the administrative data. (Of course, is is possible that city officials revised the classification of individual crimes after witnesses reported them. We cannot rule out this possibility. Still, the discrepancies between the administrative and survey data are striking.)

Other potential explanations are more speculative, and are difficult to test with our data. For example, it is possible that crime is a function of persistent socioeconomic conditions that military policing cannot address. In this case, a more promising deterrence strategy might involve social welfare programs for citizens or desistance interventions for convicted criminals [48]. The literature on these initiatives is vast and beyond the scope of our study (see [49] for a review). Whatever the explanation, our finding that Plan Fortaleza may have actually exacerbated crime belies one of the key purported benefits of military policing.

Skeptics of military policing 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 some evidence of increased abuses in surveys administered to residents. 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. In this way, military policing in Cali may have created additional opportunities for police officers to commit abuse.

Our study is not without limitations. We highlight three. First, as discussed above, we evaluate the impact of military patrols in the presence of civilian monitors. Awareness of the monitors' presence may have caused soldiers to change their behavior, inducing Hawthorne effects. We discuss our approach to minimizing Hawthorne effects in Section 4.4; we believe they are unlikely to explain our results. Second, the Plan Fortaleza program was relatively short, and was confined to weekday nights. As we note in Section 4.2, most military policing interventions are similarly short. Still, we cannot be show how our results might have changed if the program had lasted longer, or if soldiers had patrolled during the day or on weekends.

Third, like many if not most experimental evaluations of place-based policing in urban settings [10, 40, 50–53], our study focuses on a single city in a single country. As noted above, Cali is similar in important ways to other Latin American cities, and Plan Fortaleza is similar to military policing interventions in other settings. Nonetheless, organized criminal groups in Cali today tend to focus on micro-territories, and gangs are incapable of exerting monopolistic (or even duopolistic) control—a significant change from several decades ago, when the notorious Cali Cartel controlled much of the Colombian drug trade. We cannot be sure how our results might generalize to settings in which organized criminal syndicates control vast swaths of territory or large segments of the black market for drugs, weapons, and other illicit goods.

With these caveats in mind, combined with similarly disappointing results from several recent observational analyses [2, 24, 27], our findings suggest that the costs of military policing likely outweigh the benefits. Rather than outsource law enforcement to institutions that were designed for other purposes, Latin American policymakers should focus on reducing police corruption, improving police training, punishing police collusion with organized crime, and increasing funding for severely resource constrained police forces. If policymakers insist on adopting military policing strategies despite the small but growing body of evidence of their ineffectiveness, they should at least complement those strategies with robust systems for monitoring and prosecuting misconduct. It may be possible to convince both militaries and police forces that such systems are in their interest; research on body-worn cameras (BWCs), for example, has found that police officers sometimes welcome BWCs to discourage “frivolous, malicious, or unfounded” complaints by citizens [54, p. 100]. More robust monitoring systems may help reduce baseless accusations of abuse, but may also help prevent abuses from occurring in the first place.

# 4 METHODS

## 4.1 ETHICS

This study was approved by the Ethics Committee at Universidad de los Andes (Acta 1073 de 2019 and Acta 1034 de 2019). Given (1) the increasing prevalence of military policing in Colombia and many other low and middle income countries, (2) the absence of evidence on its efficacy, and (3) 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 needed to inform both scholarship and 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 communes and neighborhoods in which the intervention would occur; we randomized only the specific city 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 minimize. First, there was a risk that military patrols would subject residents to human rights abuses by soldiers. To address this risk, we used the firsthand observations of civilian monitors to document any abuses as they occurred. This gave us the ability to discontinue the evaluation if we determined that military patrols were increasing the prevalence of abuses by soldiers. We also 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 above, 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 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 research staff, field supervisors, and civil society representatives to diagnose whether particular blocks posed especially acute security concerns, and we adjusted our data collection procedures 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 mitigate this risk, we recruited monitors who did not live in the two communes 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 their patrol vehicle at all times. As additional precautions, monitors 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.

More broadly, we believe the evaluation is ethically justifiable for several reasons. First, military policing is a reasonable and ethically permissible policy for the Colombian government to pursue in our study communities: given the checkered history of the Colombian police, there were sound reasons to believe that military policing might deter criminals and reduce human rights abuses, especially relative to the status quo involving the police alone. Second and related, there is no consensus among social scientists as to whether patrolling by soldiers is superior to patrolling by police officers. Where there is equipoise of this sort, randomization is justifiable as a way to allocate access to programs [55, 56].

Finally, while citizens who saw or interacted with the soldiers while on patrol did not have an opportunity to consent to participate in the study, ethicists argue that, in the context of experiments involving government policy, researchers need not secure informed consent when three conditions are met [57]: (1) the government has a "right to rule" over the policy sphere studied; (2) data collection does not violate people's autonomy rights; and (3) there is a strong justification for not securing consent. We believe this evaluation satisfies these three conditions, given that (1) the Colombian government has the right to decide which types of law enforcement interventions to implement; (2) much of the data used in the evaluation is administrative and, in the case of survey data, was collected voluntarily; and, finally, (3) the evaluation could not have been conducted if it had been necessary to obtain informed consent from every citizen who might see or interact with the soldiers while on patrol. Informed consent was, however, obtained from all respondents in the monitoring and endline surveys; the requirement for informed consent for the surveys was not waived by the Ethics Committee at Universidad de los Andes. Respondents were not compensated for their participation in the surveys. The Ethics Committee did not require informed consent from all citizens who might see or interact with the soldiers while on patrol.

### 4.2 The Plan Fortaleza program

The Plan Fortaleza program consisted of recurring, intensive vehicular and foot patrols by heavily armed soldiers from two units of the Colombian Armed Forces: 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. 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. These are times when crime spikes, and when most citizens (criminals and otherwise) are awake and either in the street or in their homes, maximizing the probability of observing soldiers on patrol. All blocks also had some police presence.

The city of Cali comprises 22 communes in total. Communes are the highest level administrative unit in the city. Plan Fortaleza focused on communes 18 and 20, both hot spots for crime, as we show in Supplementary Information A. The two communes comprise 30 "barrios" (neighborhoods). Their combined population was approximately 215,000 at the time of our study—roughly the same as Birmingham, Alabama or Reno, Nevada. To minimize logistical problems, the two units of the military never patrolled the same commune on the same day; instead, they alternated following a 12-day rotation schedule, illustrated in Supplementary Information A. 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.

Our unit of randomization is the 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 patrol day, due in part to the small size of most blocks and the large number of soldiers on patrol. (The average perimeter of blocks in our sample is 283 meters, with a standard deviation of 248 meters.) 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.

While the intervention was relatively short, it was not atypical of the way Latin American militaries often engage 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 [14, 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 [2]. 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.

Communes 18 and 20 are densely populated and difficult to navigate. In some parts of commune 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 Supplementary Information A. 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 statistics on the patrols in Supplementary Information A, and discuss our safety protocols and the possibility of Hawthorne effects induced by the monitors' presence in further detail below.

### 4.3 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. 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 Supplementary Information 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 Supplementary Information C and balance tests in Supplementary Information D. In Supplementary Information

E we explore different ways of modeling spillover effects, including saturation and linear and exponential decay. Given the close proximity of the blocks in our sample, these analyses are generally not informative, as the distances along which treatment effects could plausibly decay are very short.

## 4.4 Data

We collected data from four sources. First, we collected timestamped, geocoded administrative data on crime, including homicides, armed robberies, thefts, 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 regularly with the Colombian National Police, the Attorney-General's Office, and the Coroner's Office to approximate the "true" prevalence of homicides and other violent crimes. We also collected timestamped, geocoded administrative data on human rights abuses from the Office of the Attorney-General, which is responsible for 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 communes 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. Monitoring survey respondents were 32.1% male and 68.9% female, with an average age of 48.0 years. Respondents were not compensated for participating in the monitoring survey. We obtained informed consent from all monitoring survey respondents.

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. We refer to this as our endline survey. On five blocks (0.3% of the sample) we were unable to implement the endline survey due to safety concerns. We drop these blocks from the sample for purposes of analysis. All surveys were conducted by trained Colombian enumerators who were blinded to the treatment status of the blocks they surveyed. Endline survey respondents were 33.4% male and 66.6% female, with an average age of 46.4 years. Respondents were not compensated for participating in the endline survey. We again obtained informed consent from all endline survey respondents.

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—even without the monitors' presence, thus blunting any Hawthorne effects induced by the monitors themselves. Nonetheless, our results should be interpreted as capturing the effects of military patrols in the presence of monitors.

## 4.5 ESTIMATION

Some of our outcomes are measured at the individual level using survey data, others at the block level using administrative data. Following our PAP, we estimate the ITT 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. In addition, 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. We exclude three blocks with 0 probability of assignment to control when estimating the ITTs [58]. When estimating spillover effects, we exclude three blocks with 0 probability of assignment to control as well as one block with 0 probability of assignment to spillover.

When testing treatment and spillover effects at the block level, we estimate a weighted least squares regression given by

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

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 a vector of block-level covariates; $\alpha_k$ denotes neighborhood fixed effects; and $\epsilon_j$ is a block-level error term. Following our PAP, we control for area of the block, number of buildings on the block, 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.

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

$$y_{ijk} = \theta t_{jk} + \lambda s_{jk} + \beta \mathbf{X_{jk}} + \delta \mathbf{Z_{ijk}} + \alpha_k + \epsilon_{ijk} \tag{2}$$

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; $\alpha_k$ denotes neighborhood fixed effects; and $\epsilon_{ijk}$ is an individual-level error term, clustered by block. Again following our PAP, we control for age, gender, and—when available—years of education. (We measured years of education in the endline but not the monitoring survey.) All tests are two-tailed. Data distributions were assumed to be normal but this was not formally tested. We report results with multiple comparisons corrections in Supplementary Information F, and heterogeneous treatment effects by gender and baseline crime rates in Supplementary Information I. We use Stata 17 and R version 4.2.0 for all statistical tests.

### 4.6 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 [7, 30, 59, 60]. In equations (1) and (2) we assume that any block that is adjacent to a treatment block is susceptible to spillover, and that any block that is not adjacent is not. In Supplementary Information E we test for the possibility of more complex spillover dynamics, including linear decay, exponential decay, and saturation, though again, given the close proximity of blocks in our sample, we view these analyses as mostly uninformative.

A related concern is the fact that spillover in our case is a function of geographical proximity to treated blocks, creating dependencies in the probability of treatment assignment that span administrative units. This has been referred to as a problem of "fuzzy clustering" [40], and it is common to virtually all studies of hot spots policing (and, indeed, to most studies in which geographical spillover effects are at least plausible). Despite its pervasiveness, the problem has received scant attention in the hot spots policing literature, and methodologists continue to disagree about how and under what conditions to adjust for clustering of this sort [**abadie2023when**, 58]. We follow [40] and report randomization inference (RI) $p$-values alongside the more conventional $p$-values and confidence intervals for all our analyses. RI tests the sharp null hypothesis of no treatment effect for any unit. While this approach to hypothesis testing remains controversial [61, 62], RI $p$-values provide a useful (and generally more conservative) complement to more conventional test statistics in the presence of fuzzy clustering.

## 4.7 Treatment 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 Supplementary Information J.1 we show that residents of treatment blocks were more likely to report military presence than residents of control blocks during the intervention ($\beta = 0.105$, CI $= [0.034, 0.175]$, $p = 0.004$, RI $p = 0.064$). We find little to no credible evidence that residents of treatment blocks were more likely to report police presence ($\beta = 0.032$, CI $= [-0.049, 0.112]$, $p = 0.439$, RI $p = 0.611$, $B_{10} = 0.067$). This is as expected, and is indicative of treatment compliance.

## 4.8 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 Supplementary Information G. 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 homicides, assaults, thefts, car thefts, and motorcycle thefts. We were in fact able to collect administrative crime data on homicides, robberies (including armed robbery and all types of theft), illegal drug sales, and illegal possession of weapons. To avoid discarding potentially useful data, we include all 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 these data from the government of Cali, and 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 [63]. 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.

## PROTOCOL REGISTRATION

This study was registered with the Evidence in Governance and Politics (EGAP) network prior to endline data collection. Our pre-analysis plan (PAP) is available at https://osf.io/95cz3?mode=&revisionId=&view_only=.

## DATA AVAILABILITY STATEMENT

All data generated or analysed for this study can be found on Dataverse at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/WAJ9SR.

## CODE AVAILABILITY STATEMENT

All code required to replicate the analyses in this study can be found on Dataverse at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/WAJ9SR.

## ACKNOWLEDGEMENTS

Funding for this project was provided in part by the UK Foreign, Commonwealth, & Development Office, awarded through Innovation for Poverty Action's Peace & Recovery Program (grant #MIT0019-S3). Funding for this project was also provided in part by the Open Society Foundations, awarded through Innovation for Poverty Action's Peace & Recovery Program (grant #OSF-19-10002-S1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

We thank staff at IPA Colombia (Kyle Holloway and Sofía Jaramillo) for coordination of the project; the Secretaría de Seguridad de Cali (Mará Alejandra Arboleda, Pablo Gracia, Alberto Sánchez, Juan Diego Tabares, Pablo Uribe, and Andrés Villamizar) for logistical oversight; the Third Battalion of the Armed Forces of Colombia (especially Colonel Omar Arciniegas) for its commitment to the evaluation; and Carlos Bohm, María Fernanda Cortés, Abraham Farfán, Sergio Flórez, Carlos Gúzman, Sebastián Hernández, and Lucía Mendoza Mora for terrific research assistance.

For helpful feedback we thank Peter Aronow, Graeme Blair, Chris Blattman, Alex Coppock, Erica De Bruin, Leopoldo Fergusson, Don Green, Andrés Moya, Molly Offer-Westort, Matt Ross, Sarah Parkinson, Santiago Tobón, Juan Vargas, Jess Zarkin, and participants at workshops and conferences convened by the American Political Science Association, the American Economic Association, Brown University, Cornell University, Duke University, the Folke Bernadotte Academy, Harvard University's Belfer Center, the Households in Conflict Network, the International Studies Association, Oxford University, Peace Research Institute Oslo, Universidad de los Andes, Universidad del Valle, University of California–Berkeley, University of Pennsylvania, Uppsala University, and the Virtual Crime Economics Online Seminar.

## AUTHOR CONTRIBUTIONS STATEMENT

RB and MW contributed equally to this article.

## COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

Table 1: Randomized military patrols had little to no effect on crime while the intervention was ongoing, and had adverse effects after the intervention was complete

| | **Admin data** | | **Endline survey** | | |
|---|---|---|---|---|---|
| | **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 |
| | [-0.068, 0.074] | [0.011, 0.208] | [-0.077, 0.089] | [-0.098, 0.085] | [0.051, 0.256] |
| | (0.934) | (0.029) | (0.886) | (0.886) | (0.003) |
| Spillover | -0.038 | 0.083 | 0.026 | 0.013 | 0.186 |
| | [-0.097, 0.022] | [-0.003, 0.169] | [-0.034, 0.086] | [-0.061, 0.087] | [0.101, 0.270] |
| | (0.212) | (0.059) | (0.389) | (0.729) | (<0.001) |
| Individual controls | No | No | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes | Yes | Yes |
| 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 |
| RI $p$-value (treatment) | 0.959 | 0.136 | 0.927 | 0.914 | 0.038 |
| RI $p$-value (spillover) | 0.411 | 0.138 | 0.610 | 0.802 | 0.001 |

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. The dependent variable in columns 1 and 2 is an additive index of murders, armed robberies, thefts, illegal drug sales, and illegal possession of weapons. The dependent variable in columns 3 and 4 is a standardized additive index of victimization by vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, domestic violence, or extortion. The dependent variable in column 5 is a standardized additive index of witnessing vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, domestic violence, extortion, prostitution, illegal drug sales, illegal drug consumption, public alcohol consumption, or illegal possession of firearms in the past month. The frequency of crime witnessing was measured on a 1-4 Likert scale. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. Models 1 and 2 also include a lagged dependent variable and block-level controls for the average age, average years of education, and percentage of men on each block. Models 3-5 include individual-level controls for age, gender, and years of education. ITT effect sizes in columns 1 and 2 are derived from WLS regressions as in equation 1. ITT effect sizes in columns 3-5 are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors in models 3-6 are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. Randomization inference $p$-values are in the last two rows. No adjustments were made for multiple comparisons.

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

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

Figure 2: Randomized military patrols had little to no effect on crime even while soldiers were physically present on the streets

Notes: ITT on crime during the intervention based on administrative data, disaggregated by day and time. The dependent variable in all models is an additive index of murders, armed robberies, thefts, illegal drug sales, and illegal possession of weapons. All models include neighborhood fixed effects, a lagged dependent variable, and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include block-level controls for the average age, average years of education, and percentage of men on each block. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "all crime" ($N = 1,167$): $\beta = 0.003$, CI = [-0.068, 0.074], $p = 0.934$, RI $p = 0.957$ on treatment blocks; $\beta = -0.038$, CI = [-0.097, 0.022], $p = 0.212$, RI $p = 0.413$ on spillover blocks. For "weekend crime" ($N = 1,167$): $\beta = -0.003$, CI = [-0.035, 0.029], $p = 0.857$, RI $p = 0.890$ on treatment blocks; $\beta = 0.001$, CI = [-0.028, 0.030], $p = 0.961$, RI $p = 0.968$ on spillover blocks. For "weekday crime" ($N = 1,167$): $\beta = 0.006$, CI = [-0.053, 0.066], $p = 0.834$, RI $p = 0.901$ on treatment blocks; $\beta = -0.038$, CI = [-0.088, 0.011], $p = 0.129$, RI $p = 0.365$ on spillover blocks. For "daytime crime" ($N = 1,167$): $\beta = 0.011$, CI = [-0.042, 0.064], $p = 0.681$, RI $p = 0.786$ on treatment blocks; $\beta = -0.022$, CI = [-0.065, 0.022], $p = 0.330$, RI $p = 0.510$ on spillover blocks. For "nighttime crime" ($N = 1,167$): $\beta = -0.008$, CI = [-0.054, 0.038], $p = 0.741$, RI $p = 0.812$ on treatment blocks; $\beta = -0.017$, CI = [-0.060, 0.025], $p = 0.424$, RI $p = 0.514$ on spillover blocks.

Figure 3: Randomized military patrols did not improve perceptions of safety, except among business owners

Notes: ITT on perceptions of safety for residents and business owners together (“all safety”), residents only (“personal safety”), and business owners only (“business safety”) based on survey data. Residents and business owners were asked how safe they feel walking their blocks during the day and at night; how safe they feel talking on a smartphone on their blowk; and how worried they are about becoming victims of a violent or non-violent crime in the next two weeks. Perceptions of safety were measured on a 1-5 Likert scale. Residents and business owners were also asked whether they had taken precautions for fear of crime in the past month, including avoiding leaving their home or business at night; avoiding public transportation or recreation areas; prohibiting children from playing in the streets or attending school; considering moving to a different neighborhood; or changing jobs or schools. The dependent variable in models 1 (“all safety”) and 2 (“personal safety”) is a standardized additive index based on responses to these questions. Business owners were also asked if they had changed their hours, hired private security guards, or closed their businesses for fear of crime. The dependent variable in model 3 (“business safety”) is a standardized additive index based on responses to these questions. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons; results with multiple comparisons corrections are reported in Tables 8 and 9. For “all safety” ($N = 7,707$): $\beta = -0.050$, CI = [-0.142, 0.043], $p = 0.292$, RI $p = 0.449$ on treatment blocks; $\beta = -0.066$, CI = [-0.144, 0.013], $p = 0.100$, RI $p = 0.200$ on spillover blocks. For “personal safety” ($N = 7,708$): $\beta = -0.052$, CI = [-0.144, 0.041], $p = 0.271$, RI $p = 0.431$ on treatment blocks; $\beta = -0.068$, CI = [-0.146, 0.010], $p = 0.089$, RI $p = 0.186$ on spillover blocks. For “business safety” ($N = 1,014$): $\beta = 0.284$, CI = [0.056, 0.512], $p = 0.015$, RI $p = 0.064$ on treatment blocks; $\beta = 0.094$, CI = [-0.110, 0.297], $p = 0.367$, RI $p = 0.472$ on spillover blocks.

Figure 4: Randomized military patrols may have exacerbated human rights abuses, especially by the police

Notes: ITT on abuses committed by police officers and soldiers based on monitoring (top panel) and survey data (bottom panel). The dependent variable in all models is an indicator for any incidents of verbal or physical abuse on the block. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age and gender; models in the bottom panel include an additional individual-level control for years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons; results with multiple comparisons corrections are reported in Tables 10 and 11. For "police abuse during intervention" ($N = 1,970$): $\beta = 0.037$, CI = [0.016, 0.058], $p = 0.001$, RI $p = 0.103$ on treatment blocks; $\beta = 0.016$, CI = [-0.005, 0.037], $p = 0.137$, RI $p = 0.385$ on spillover blocks. For "military abuse during intervention" ($N = 1,970$): $\beta = 0.010$, CI = [-0.001, 0.022], $p = 0.071$, RI $p = 0.430$ on treatment blocks; $\beta = 0.001$, CI = [-0.005, 0.007], $p = 0.778$, RI $p = 0.942$ on spillover blocks. For "police abuse after intervention" ($N = 7,908$): $\beta = 0.011$, CI = [-0.015, 0.037], $p = 0.417$, RI $p = 0.541$ on treatment blocks; $\beta = 0.030$, CI = [0.007, 0.053], $p = 0.011$, RI $p = 0.028$ on spillover blocks. For "military abuse after intervention" ($N = 7,908$): $\beta = -0.001$, CI = [-0.009, 0.006], $p = 0.716$, RI $p = 0.783$ on treatment blocks; $\beta = 0.002$, CI = [-0.005, 0.009], $p = 0.544$, RI $p = 0.613$ on spillover blocks.

# REFERENCES


1. Collazos, D., García, E., Mejía, D., Ortega, D. & Tobón, S. Hot Spots Policing in a High Crime Environment: An Experimental Evaluation in Medellín. *Journal of Experimental Criminology* **17,** 473–506 (2021).
2. 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.
3. Meliala, A. Police as Military: Indonesia's Experience. *Policing: An International Journal of Police Strategies & Management* **24,** 420–432 (2001).
4. Varona, G. Politics and Policing in the Philippines: Challenges to Police Reform. *Flinders Journal of History and Politics* **26,** 102–125 (2010).
5. Montesh, M. & Basdeo, V. The Role of the South African National Defence Force in Policing. *Scientia Militaria: South African Journal of Military Studies* **40,** 71–94 (2012).
6. Dunlap Jr., C. J. The Police-Ization of the Military. *Journal of Political and Military Sociology* **27,** 397–418 (1999).
7. Braga, A. A. & Weisburd, D. L. *Policing Problem Places: Crime Hot Spots and Effective Prevention* (Oxford University Press, New York, 2010).
8. Braga, A. A., Andresen, M. A. & Lawton, B. The Law of Crime Concentration at Places: Editors' Introduction. *Journal of Quantitative Criminology* **33,** 421–426 (2017).
9. Sherman, L. W., Gartin, P. R. & Buerger, M. E. Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place. *Criminology* **27,** 27–56 (1989).
10. Sherman, L. W. & Weisburd, D. General Deterrent Effects of Police Patrol in Crime 'Hot Spots': A Randomized, Controlled Trial. *Justice Quarterly* **12,** 625–648 (1995).

11. Weisburd, D. The Law of Crime Concentration and the Criminology of Place. *Criminology* **53,** 133–157 (2015).

12. Nagin, D. in *Contemporary Issues in Criminological Theory and Research: The Role of Social Institutions (Papers from the American Society of Criminology 2010 Conference)* (eds Rosenfeld, R., Quinet, K. & Garcia, C. A.) 309–316 (Wadsworth, Belmont, CA, 2010).

13. Nagin, D. S., Solow, R. M. & Lum, C. Deterrence, Criminal Opportunities, and Police. *Criminology* **53,** 74–100 (2015).

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

15. Pion-Berlin, D. A Tale of Two Missions: Mexican Military Police Patrols Versus High-Value Targeted Operations. *Armed Forces & Society* **43,** 53–71 (2017).

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

17. Apel, R. & Nagin, D. S. in *Crime and Public Policy* (eds Wilson, J. Q. & Petersilia, J.) Revised (Oxford University Press, Oxford, 2011).

18. Gibbs, J. P. *Crime, Punishment, and Deterrence* (Elsevier, New York, 1975).

19. Kleiman, M. A. R. *When Brute Force Fails: How to Have Less Crime and Less Punishment* Reprint (Princeton University Press, Princeton, NJ, 2010).

20. Nagin, D. S. in *Crime and Justice: A Review of Research* (ed Tonry, M.) 199–263 (University of Chicago Press, Chicago, 2013).

21. Paternoster, R. The Deterrent Effect of the Perceived Certainty and Severity of Punishment: A Review of the Evidence and Issues. *Justice Quarterly* **4,** 173–217 (1987).

22. 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).

23. Pion-Berlin, D. & Carreras, M. Armed Forces, Police and Crime-Fighting in Latin America. *Journal of Politics in Latin America* **9,** 3–26 (2017).

24. 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).

25. Bayley, D. H. *What Works in Policing* (Oxford University Press, New York, 1998).

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

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

28. 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).

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

30. 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).

31. 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).

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

33. 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).

34. 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).

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

36. 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).

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

38. 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).

39. 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).

40. Blattman, C., Green, D. P., Ortega, D. & Tobón, S. Place-Based Interventions at Scale: The Direct and Spillover Effects of Policing and City Services on Crime. *Journal of the European Economic Association* **19,** 2022–2051 (2021).

41. Jeffreys, H. *Theory of Probability* Third (Oxford University Press, Oxford, 1961).

42. Lee, M. D. & Wagenmakers, E.-J. *Bayesian Cognitive Modeling: A Practical Course* (Cambridge University Press, Cambridge, UK, 2014).

43. Quintana, D. S. & Williams, D. R. Bayesian Alternatives for Common Null-Hypothesis Significance Tests in Psychiatry: A Non-Technical Guide Using JASP. *BMC Psychiatry* **18,** 178 (2018).

44. Kochel, T. R. Constructing Hot Spots Policing: Unexamined Consequences for Disadvantaged Populations and for Police Legitimacy. *Criminal Justice Policy Review* **22,** 350–374 (2011).

45. Rosenbaum, D. P. in *Police Innovation: Contrasting Perspectives* (eds Weisburd, D. & Braga, A. A.) 245–266 (Cambridge University Press, Cambridge, UK, 2006).

46. Weisburd, D., Hinkle, J. C., Famega, C. & Ready, J. The Possible "Backfire" Effects of Hot Spots Policing: An Experimental Assessment of Impacts on Legitimacy, Fear and Collective Efficacy. *Journal of Experimental Criminology* **7,** 297–320 (2011).

47. Ratcliffe, J. H., Groff, E. R., Sorg, E. T. & Haberman, C. P. Citizens' Reactions to Hot Spots Policing: Impacts on Perceptions of Crime, Disorder, Safety and Police. *Journal of Experimental Criminology* **11,** 393–417 (2015).

48. Carr, J. B. & Packham, A. SNAP Benefits and Crime: Evidence from Changing Disbursement Schedules. *The Review of Economics and Statistics* **101,** 310–325 (2019).

49. Doleac, J. L. *Encouraging Desistance from Crime* SSRN Scholarly Paper 3825106. 2020. https://papers.ssrn.com/abstract=3825106.

50. Braga, A. A. & Bond, B. J. Policing Crime and Disorder Hot Spots: A Randomized Controlled Trial. *Criminology* **46,** 577–607 (2008).

51. Corsaro, N., Brunson, R. K. & McGarrell, E. F. Evaluating a policing strategy intended to disrupt an illicit street-level drug market. *Evaluation review* **34,** 513–548 (2010).

52. Ratcliffe, J. H., Taniguchi, T., Groff, E. R. & Wood, J. D. The Philadelphia Foot Patrol Experiment: A Randomized Controlled Trial of Police Patrol Effectiveness in Violent Crime Hotspots. *Criminology* **49,** 795–831 (2011).

53. Weisburd, D. & Green, L. Policing Drug Hot Spots: The Jersey City Drug Market Analysis Experiment. *Justice Quarterly* **12,** 711–735 (1995).

54. Lum, C., Stoltz, M., Koper, C. S. & Scherer, J. A. Research on Body-Worn Cameras: What We Know, What We Need to Know. *Criminology & Public Policy* **18,** 93–118 (2019).

55. MacKay, D. The ethics of public policy RCTs: The principle of policy equipoise. *Bioethics* **32,** 59–67 (2018).

56. Asiedu, E., Karlan, D., Lambon-Quayefio, M. P. & Udry, C. R. *A Call for structured ethics Appendices in social science papers* NBER Working Paper 28393. 2021. https://www.nber.org/papers/w28393.

57. MacKay, D. & Chakrabarti, A. Government policy experiments and informed consent. *Public Health Ethics* **12,** 188–201 (2019).

58. 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).

59. 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).

60. Weisburd, D. *et al.* Does Crime Just Move Around the Corner? A Controlled Study of Spatial Displacement and Diffusion of Crime Control Benefits. *Criminology* **44,** 549–592 (2006).

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

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

63. Aronow, P., Samii, C. D. & Wang, Y. *Design-Based Inference for Spatial Experiments with Interference* Working paper. 2019. https://arxiv.org/abs/2010.13599.

64. 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).

# “Little evidence that military policing reduces crime or improves human security”

## Supplementary Information

# A Setting and implementation

Supplementary Table 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 commune on the same day. Supplementary Figures 1 and 2 show screenshots from the customized Google Maps interface used by civilian monitors to help the soldiers navigate to treatment blocks, and to record the soldiers' activities while on patrol. Supplementary Table 2 provides descriptive statistics on the soldiers' activities while on patrol based on the monitors' reports. Supplementary Figure 3 plots the distribution of homicides across the 22 communes 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. Supplementary Figures 4 and 5 map the blocks in our sample by treatment assignment.

# C Power calculations

We estimate our minimum detectable effects (MDEs) using administrative crime data from before the start of the intervention. We first bootstrap our randomization procedure 1,000 times. Within each replication we simulate the treatment and spillover effects of the intervention, knowing that the true effect must be 0. The dependent variable is identical to the one we use in the paper, except that it is standardized for ease of interpretation, and only includes crimes committed during the four-month period between July 1 and September 29, 2019—the day before the intervention began. All replications include neighborhood fixed effects, inverse probability weights, and a lagged dependent variable for the number of crimes committed on each block between January 1 and June 30, 2019.

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

$$y_{jk} = \theta t_{jk} + \lambda s_{jk} + \gamma y_{jk,t-1} + \alpha_k + \epsilon_{jk} \tag{SI.1}$$

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

Without covariates we find that we are powered to detect treatment effects of approximately 0.18 standard deviations and spillover effects of approximately 0.14 standard deviations.[1] These would generally be considered small MDEs, meaning that we should be able to detect even substantively modest reductions in crime in our analyses in the paper. These MDEs are also likely conservative, since they do not include covariates, and since the dependent variable is operationalized at the block level. Our inclusion of covariates and our use of survey data (which increases our

[1]Our MDE estimates are somewhat sensitive to the temporal window we use to define the dependent variable: they become larger as the temporal window expands, and narrower as it contracts. For example, if we define the dependent variable to include crimes committed between June 1 and September 29, we are powered to detect treatment effects of approximately 0.21 standard deviations and spillover effects of approximately 0.17 standard deviations. If we instead define the dependent variable to include crimes committed between May 1 and September 29, we are powered to detect treatment effects of approximately 0.22 standard deviations and spillover effects of approximately 0.17 standard deviations. The temporal window of the dependent variables in the paper is narrower than any of those described above (seven weeks for the period during the intervention, six weeks for the period after), again suggesting that our MDE estimates are likely conservative.

sample size) for most analyses in the paper should further improve our statistical power.

# D Balance tests

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

# E Spillover

In the paper we assume that treatment effects can only spill over from treated blocks to adjacent control blocks. In Supplementary Table 5 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

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

where $f(d_{jk})$ is a linear decay function with a standardized distribution and $f(d_{jk}) = \frac{1}{d_{jk}}$. This function is a weighted sum of distances to all 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 assume that treatment effects become weaker as an exponential function of distance to the nearest treated block. We estimate

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

where $g(d_{jk})$ is a spatial decay function with a standardized distribution and $g(d_{jk}) = \frac{1}{e^{d_{jk}}}$. We calculate the distance, $d$, to all 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. We estimate

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

where $y_{jk}$ denotes the outcome for block $j$ in neighborhood $k$; $p$ denotes the proportion of adjacent blocks assigned to the treatment group; $\mathbf{X_{jk}}$ denotes block-level covariates; $\alpha_k$ denotes neighborhood fixed effects; and $\epsilon_j$ is a block-level error term. All estimates are weighted by the inverse probability of assignment to each block's realized treatment status.

# F MULTIPLE COMPARISONS

In the paper we use indexing to reduce the number of hypotheses we test. In Supplementary Tables 6 through 11 we replicate all analyses and report the Benjamini-Hochberg $q$-value and Holm-Bonferroni threshold for our treatment and spillover indicators. Following [64], 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. Supplementary Tables 6 and 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 corresponding outcome.

# G Crime

## G.1 Weighted standardized crime indices

Supplementary Figure 6 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. Supplementary Figure 7 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.

## G.2 Violent and non-violent crime indices

Supplementary Figure 8 reports the ITT of the Plan Fortaleza program on additive indices of violent and non-violent crimes committed during (top panel) and after (bottom panel) the intervention using administrative data. Supplementary Table 12 reports the ITT on standardized additive indices of violent crime victimization and witnessing in the survey; Supplementary Table 13 reports the ITT on standardized additive indices of non-violent crime victimization and witnessing.

## G.3 Crime disaggregated by type

Supplementary Figure 9 reports the ITT of the Plan Fortaleza program on specific categories of crime in the administrative data. Supplementary Figures 10 and 11 report the ITT on specific catetogries of crime victimization and witnessing, respectively, in the survey.

# H ABUSES

## H.1 ABUSES IN ADMINISTRATIVE DATA

Supplementary Figure 12 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 use the date the abuse was alleged to have occurred, rather than the date the report was filed. Supplementary Figures 13 and 14 focus on the two communes in our sample. Green denotes treatment blocks, yellow spillover blocks, and red control blocks.

Only one allegation of abuse was recorded during the intervention. It involved a transit police officer who was accused of unfairly restricting a citizen's freedom of movement. Two more allegations were recorded after the intervention was complete. The first occurred in commune 18 on November 21 on a spillover block, during the afternoon, 173 meters from the nearest treatment block. The witness reported that she was walking with her husband when two men ran by. A group of police officers caught the men and allegedy beat them, likely while effecting an arrest. When the witness began filming the incident, the officers allegedly attacked her and her husband, seizing both of their cell phones, presumably to destroy evidence. The second allegation occurred on December 6 on a control block in commune 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 allegedly threw rocks at him. Given the nature, location, and timing of these events, we believe it is unlikely that they were related to Plan Fortaleza in any direct way.

### H.2 Abuses in monitoring survey data

Supplementary Figures 15 through 22 plot the distribution of abuses committed by soldiers and police officers across the two communes 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.

### H.3 Abuses in monitoring survey disaggregated by period

Supplementary Figures 23 and 24 replicate our results in the top panel of Figure 4 using two different approaches to defining the period during and after the intervention. In Supplementary Figure 23 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 about abuses they witnessed or heard about in the two weeks prior to being surveyed. As an additional robustness check, in Supplementary Figure 24 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.

## I Heterogeneous treatment effects

Supplementary Table 14 reports heterogeneous treatment effects (HTEs) 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. Supplementary Table 15 reports HTEs 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 HTEs on crime in the administrative data by gender in the survey.) Supplementary Tables 16 and 17 report HTEs on perceptions of security by prior crime rate in the administrative data and gender, respectively. Supplementary Tables 18 and 19 report HTEs on abuses by prior crime rate and gender, respectively.

Supplementary Table 20 reports HTEs on crime victimization and witnessing by prior crime victimization. To operationalize prior crime victimization, we create a standardized additive index of respondents' reports of 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 HTEs on crime in the administrative data by prior crime victimization in the survey.) Finally, Supplementary Table 21 reports HTEs on perceptions of safety by prior crime victimization, while Table 22 reports HTEs on abuses by prior crime victimization. We use the same operationalization for crime victimization described above.

# J Ancillary and exploratory analyses

## J.1 Exposure to police and military

Supplementary Figure 25 reports the ITT 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 (top panel). 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 (bottom panel). This latter question does not measure treatment compliance per se, since endline survey respondents were surveyed a month or more after the program ended. But testing the ITT on this measure helps us assess whether soldiers continued patrolling the same treatment blocks even after the intervention.

### J.2 Police and military activities

Supplementary Figures 26 and 27 report the ITT of the program on military and police activities during and after the intervention, respectively. Respondents were asked how often they had seen or heard about the police or military making arrests on their block; how often they had seen or heard about the police or military checking IDs on their block; and how often they had seen or heard about the police or military talking with citizens on their block. The monitoring survey (Supplementary Figure 26) asked respondents whether they had seen or heard about these activities in the prior two weeks, while the endline survey (Supplementary Figure 27) asked respondents whether they had seen or heard about these activities in the past month. Frequency was measured on a Likert scale from 1 to 4.

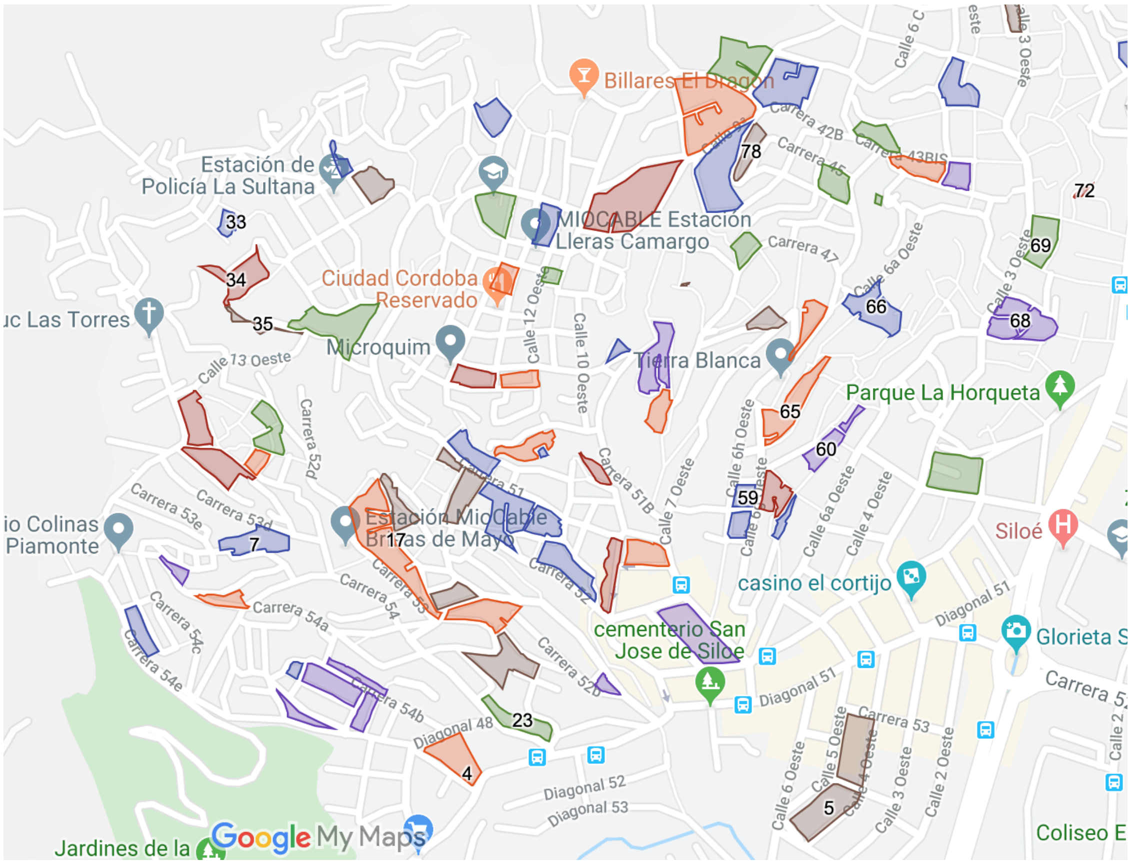
### J.3 Cooperation with police and military

Supplementary Figure 28 reports the ITT of the program on cooperation with the authorities in general (top panel), and the police (middle panel) and military (bottom panel) 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 additive indices of cooperation using responses to these questions.

Supplementary Table 1: 12-day rotation schedule for Special Forces and Military Police

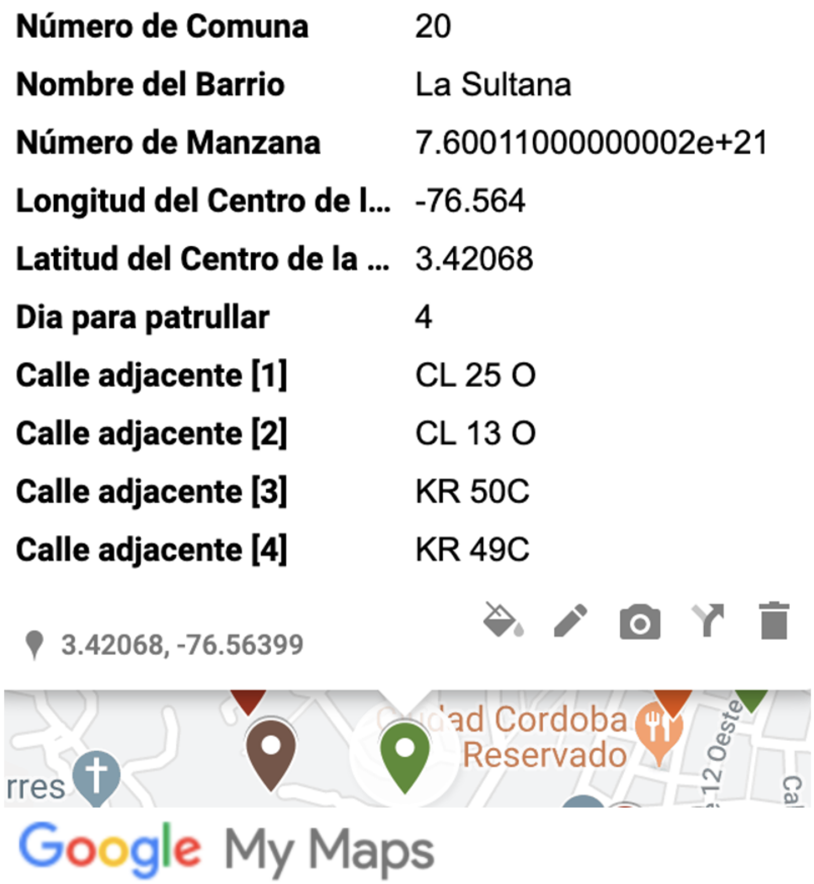
| | Commune 18 | Commune 20 |
|---|---|---|
| Day 1 | Special Forces | Military Police |
| Day 2 | Military Police | Special Forces |
| Day 3 | Special Forces | Military Police |
| Day 4 | Military Police | Special Forces |
| Day 5 | Special Forces | Military Police |
| Day 6 | Military Police | Special 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 |

Supplementary Figure 1: Screenshot of smartphone user interface with locations of treatment blocks in commune 20



Notes: Colors indicate which day a treatment block was assigned to be patrolled. Source: Google Maps (underlying map).

Supplementary Figure 2: Screenshot of smartphone user interface with information on a single treatment block



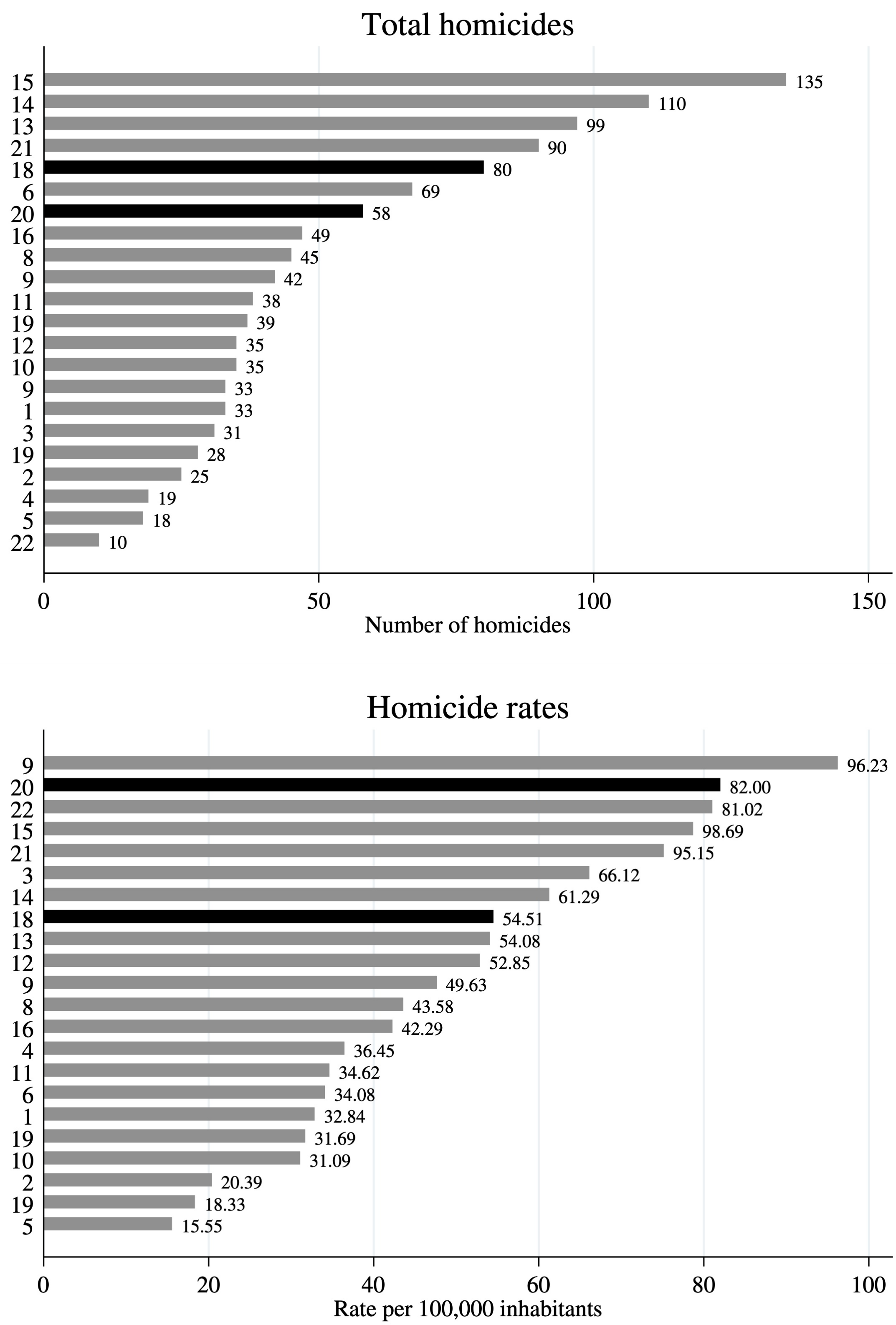

Notes: Information provided to soldiers included latitude and longitude, patrol day, and cross streets for each treatment block. Source: Google Maps (underlying map).

Supplementary Table 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 |

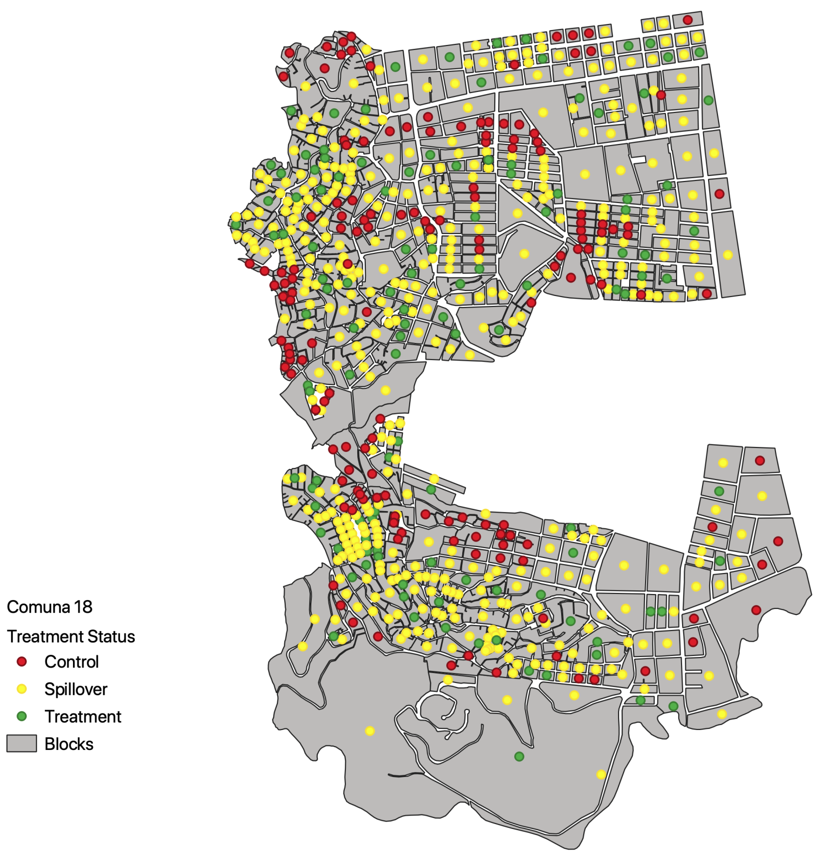
Notes: All descriptive statistics are based on GIS data and the first-hand observations of civilian monitors. Standard deviations in parentheses.

Supplementary Figure 3: Homicides in Cali by commune using administrative data



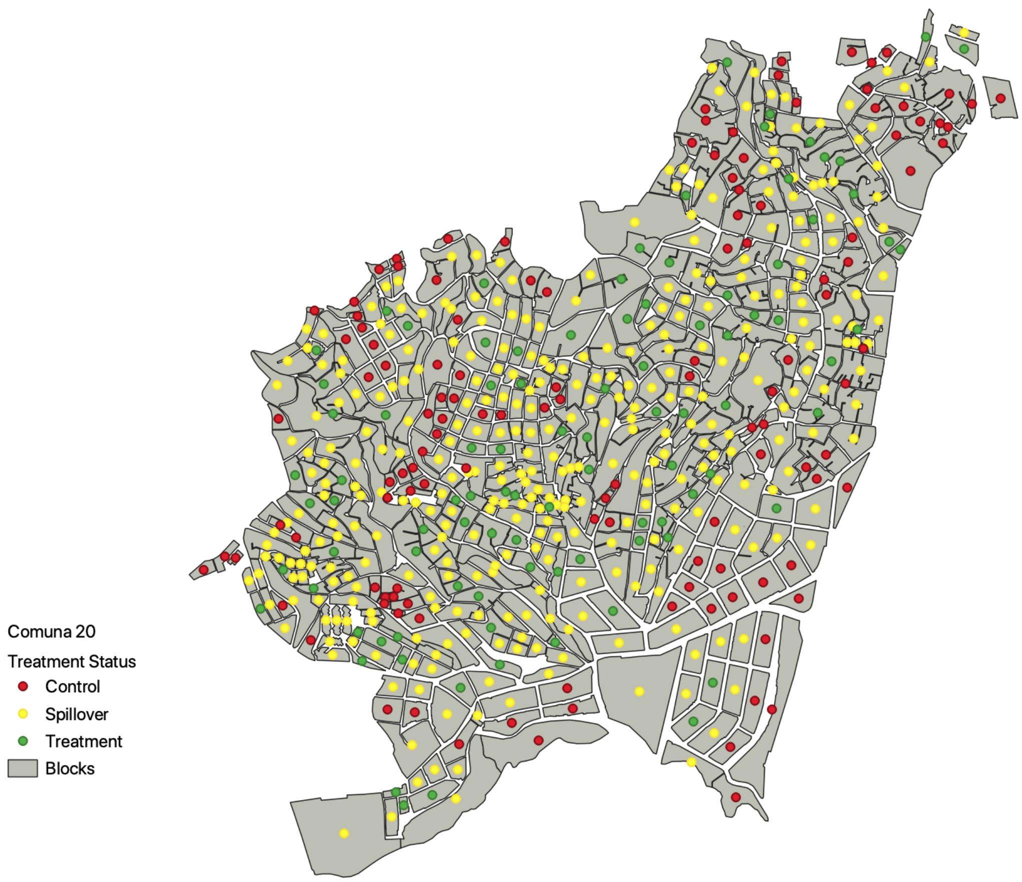

Source: Administrative crime data, Mayor's Office of Cali.

Supplementary Figure 4: Assignment to treatment in commune 18




Notes: Green denotes treatment blocks, yellow denotes spillover blocks, and red denotes control blocks. Source: Instituto Geográfico Agustín Codazzi - Igac.

Supplementary Figure 5: Assignment to treatment in commune 20




Notes: Green denotes treatment blocks, yellow denotes spillover blocks, and red denotes control blocks. Source: Instituto Geográfico Agustín Codazzi - Igac.

Supplementary Table 3: Balance using administrative data

| | Control | Treatment | Spillover |
|---|---|---|---|
| ***Panel A: Index without controls*** | | | |
| Crime index | 0.002 | 0.000 | -0.002 |
| | [-0.012 - 0.016] | [-0.020 - 0.020] | [-0.011 - 0.006] |
| | (0.756) | (0.994) | (0.599) |
| ***Panel B: Components without controls*** | | | |
| Homicides | -0.019 | -0.002 | 0.021 |
| | [-0.131 - 0.093] | [-0.132 - 0.127] | [-0.062 - 0.104] |
| | (0.742) | (0.972) | (0.618) |
| Robberies | 0.004 | -0.001 | -0.003 |
| | [-0.013 - 0.021] | [-0.023 - 0.021] | [-0.011 - 0.005] |
| | (0.633) | (0.920) | (0.465) |
| Drug dealing | -0.027 | 0.077 | -0.049 |
| | [-0.192 - 0.138] | [-0.120 - 0.274] | [-0.169 - 0.070] |
| | (0.746) | (0.445) | (0.417) |
| llegal possession of a firearm | -0.086 | 0.024 | 0.062 |
| | [-0.228 - 0.057] | [-0.224 - 0.272] | [-0.125 - 0.248] |
| | (0.238) | (0.849) | (0.517) |
| ***Panel C: Index with controls*** | | | |
| Crime index | 0.002 | 0.001 | -0.004 |
| | [-0.013 - 0.018] | [-0.019 - 0.022] | [-0.012 - 0.004] |
| | (0.767) | (0.905) | (0.385) |
| Number of buildings on block | 0.001 | -0.000 | -0.001 |
| | [-0.001 - 0.003] | [-0.003 - 0.002] | [-0.002 - 0.001] |
| | (0.256) | (0.762) | (0.515) |
| Area of block | -0.000 | -0.000 | 0.000 |
| | [-0.000 - 0.000] | [-0.000 - 0.000] | [0.000 - 0.000] |
| | (0.410) | (0.139) | (0.007) |
| Distance to nearest army battalion (meters) | 0.000 | -0.000 | -0.000 |
| | [0.000 - 0.001] | [-0.001 - 0.000] | [-0.000 - 0.000] |
| | (0.003) | (0.053) | (0.553) |
| Distance to nearest police station (meters) | 0.000 | -0.000 | -0.000 |
| | [-0.000 - 0.000] | [-0.000 - 0.000] | [-0.000 - 0.000] |
| | (0.889) | (0.947) | (0.953) |
| Distance to nearest public transportation hub (meters) | -0.000 | 0.000 | -0.000 |
| | [-0.000 - 0.000] | [-0.000 - 0.000] | [-0.000 - 0.000] |
| | (0.565) | (0.497) | (0.710) |

Notes: The dependent variables in models 1, 2, and 3 are indicators for assignment to the control, treatment, and spillover groups, respectively. The independent variable in panel A is an additive index of crime based on administrative data. The independent variables in panel B are the components of the additive crime index from panel A. The independent variables in panel C are the additive crime index and block-level controls for the area of each block; distance to the nearest police station; distance to the nearest military battalion; and distance to the nearest public transportation hub, based on administrative data. Coefficients are derived from OLS regressions. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Table 3: Balance using administrative data (cont.)

| | Control | Treatment | Spillover |
|---|---|---|---|
| ***Panel D: Components with controls*** | | | |
| Homicides | -0.023 | 0.022 | 0.001 |
| | [-0.139 - 0.093] | [-0.111 - 0.155] | [-0.084 - 0.086] |
| | (0.697) | (0.744) | (0.982) |
| Robberies | 0.004 | -0.000 | -0.004 |
| | [-0.014 - 0.022] | [-0.023 - 0.023] | [-0.012 - 0.004] |
| | (0.670) | (0.987) | (0.361) |
| Drug dealing | -0.009 | 0.060 | -0.051 |
| | [-0.182 - 0.165] | [-0.145 - 0.264] | [-0.181 - 0.079] |
| | (0.922) | (0.566) | (0.440) |
| llegal possession of a firearm | -0.089 | 0.022 | 0.067 |
| | [-0.226 - 0.047] | [-0.217 - 0.261] | [-0.126 - 0.260] |
| | (0.199) | (0.854) | (0.496) |
| Number of buildings on block | 0.001 | -0.000 | -0.001 |
| | [-0.001 - 0.003] | [-0.003 - 0.002] | [-0.002 - 0.001] |
| | (0.257) | (0.754) | (0.523) |
| Area of block | -0.000 | -0.000 | 0.000 |
| | [-0.000 - 0.000] | [-0.000 - 0.000] | [0.000 - 0.000] |
| | (0.418) | (0.133) | (0.006) |
| Distance to nearest army battalion (meters) | 0.000 | -0.000 | -0.000 |
| | [0.000 - 0.001] | [-0.001 - 0.000] | [-0.000 - 0.000] |
| | (0.003) | (0.056) | (0.563) |
| Distance to nearest police station (meters) | 0.000 | -0.000 | -0.000 |
| | [-0.000 - 0.000] | [-0.000 - 0.000] | [-0.000 - 0.000] |
| | (0.846) | (0.944) | (0.908) |
| Distance to nearest public transportation hub (meters) | -0.000 | 0.000 | -0.000 |
| | [-0.000 - 0.000] | [-0.000 - 0.000] | [-0.000 - 0.000] |
| | (0.610) | (0.530) | (0.711) |
| Observations | 1,254 | 1,254 | 1,254 |
| **Panel A:** $F$-stat on index without controls | 0.097 | 0.000 | 0.276 |
| | (0.756) | (0.994) | (0.599) |
| **Panel B:** $F$-stat on components without controls | 0.517 | 0.164 | 0.433 |
| | (0.723) | (0.956) | (0.785) |
| **Panel C:** $F$-stat on index | 0.088 | 0.014 | 0.756 |
| | (0.767) | (0.905) | (0.385) |
| **Panel C:** $F$-stat on index with controls | 1.941 | 1.218 | 1.561 |
| | (0.071) | (0.294) | (0.155) |
| **Panel D:** $F$-stat on components | 0.545 | 0.128 | 0.468 |
| | (0.702) | (0.972) | (0.759) |
| **Panel D:** $F$-stat on components with controls | 1.633 | 0.868 | 1.152 |
| | (0.101) | (0.553) | (0.323) |

Notes: The dependent variables in models 1, 2, and 3 are indicators for assignment to the control, treatment, and spillover groups, respectively. The independent variables in panel D are the components of the additive crime index from panel A and the block-level controls from panel C. Coefficients are derived from OLS regressions. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Table 4: Balance using endline survey data

| | Control | Treatment | Spillover |
|---|---|---|---|
| ***Panel A: Demographic controls*** | | | |
| Age | -0.001 | 0.001 | -0.000 |
| | [-0.002 - 0.000] | [-0.000 - 0.002] | [-0.001 - 0.000] |
| | (0.280) | (0.164) | (0.389) |
| Gender | 0.007 | -0.007 | -0.000 |
| | [-0.017 - 0.032] | [-0.038 - 0.024] | [-0.020 - 0.019] |
| | (0.556) | (0.655) | (0.978) |
| Education (years) | 0.001 | -0.001 | 0.000 |
| | [-0.003 - 0.005] | [-0.007 - 0.004] | [-0.003 - 0.003] |
| | (0.554) | (0.655) | (0.984) |
| ***Panel B: Demographic and geo controls*** | | | |
| Age | -0.000 | 0.001 | -0.001 |
| | [-0.001 - 0.001] | [-0.000 - 0.002] | [-0.001 - 0.000] |
| | (0.825) | (0.236) | (0.117) |
| Gender | 0.006 | -0.006 | 0.000 |
| | [-0.018 - 0.030] | [-0.037 - 0.025] | [-0.019 - 0.020] |
| | (0.629) | (0.682) | (0.963) |
| Education (years) | 0.003 | -0.002 | -0.001 |
| | [-0.000 - 0.007] | [-0.007 - 0.003] | [-0.004 - 0.002] |
| | (0.069) | (0.403) | (0.393) |
| Number of buildings on the block | -0.000 | 0.000 | -0.000 |
| | [-0.002 - 0.002] | [-0.002 - 0.003] | [-0.002 - 0.001] |
| | (0.929) | (0.686) | (0.632) |
| Area of block | -0.000 | -0.000 | 0.000 |
| | [-0.000 - 0.000] | [-0.000 - -0.000] | [0.000 - 0.000] |
| | (0.753) | (0.031) | (0.003) |
| Distance to nearest army battalion (meters) | 0.000 | -0.000 | -0.000 |
| | [0.000 - 0.000] | [-0.000 - 0.000] | [-0.000 - 0.000] |
| | (0.031) | (0.412) | (0.145) |
| Distance to nearest police station (meters) | -0.000 | 0.000 | 0.000 |
| | [-0.000 - 0.000] | [-0.000 - 0.000] | [-0.000 - 0.000] |
| | (0.212) | (0.836) | (0.226) |
| Distance to nearest public transportation hub (meters) | 0.000 | -0.000 | -0.000 |
| | [0.000 - 0.000] | [-0.000 - 0.000] | [-0.000 - 0.000] |
| | (0.045) | (0.271) | (0.517) |
| Observations | 7,918 | 7,918 | 7,918 |
| **Panel A**: F-stat individual-level controls | 1.484 | 1.708 | 0.462 |
| | (0.217) | (0.164) | (0.709) |
| **Panel B**: F-stat on block-level controls | 2.060 | 1.680 | 0.824 |
| | (0.104) | (0.169) | (0.480) |
| **Panel B**: F-stat on individual and block-level controls | 1.609 | 1.321 | 1.867 |
| | (0.118) | (0.229) | (0.062) |

Notes: The dependent variables in models 1, 2, and 3 are indicators for assignment to the control, treatment, and spillover groups, respectively. The independent variables in panel A are individual-level controls for age, gender, and years of education. The independent variables in panel B are the individual-level controls from panel A and block-level controls for the area of each block; distance to the nearest police station; distance to the nearest military battalion; and distance to the nearest public transportation hub, based on administrative data. Coefficients are derived from OLS regressions. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Table 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 |
| | [-0.037, 0.083] | [-0.018, 0.141] | [-0.037, 0.083] | [-0.021, 0.139] | [-0.341, 0.408] | [-1.010, 0.225] |
| | (0.458) | (0.129) | (0.452) | (0.150) | (0.860) | (0.211) |
| Spillover | -0.007 | -0.070 | -0.000 | -0.096 | | |
| | [-0.042, 0.029] | [-0.136, -0.005] | [-0.078, 0.078] | [-0.190, -0.001] | | |
| | (0.712) | (0.036) | (0.998) | (0.047) | | |
| Individual controls | No | No | No | No | No | No |
| Neighborhood FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes | Yes | Yes | Yes |
| 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 approaches to estimating spillover effects. The dependent variable is an additive index of murders, armed robberies, thefts, illegal drug sales, and illegal possession of weapons. All models include neighborhood fixed effects, a lagged dependent variable, and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include block-level controls for the average age, average years of education, and percentage of men on each block. ITT effect sizes in columns 1 and 2 are derived from WLS regressions as in equation SI.2. ITT effect sizes in columns 3 and 4 are derived from WLS regressions as in equation SI.3. ITT effect sizes in columns 5 and 6 are derived from WLS regressions as in equation SI.4. Observations are weighted by the inverse probability of assignment to their realized treatment status. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Table 6: Treatment effects on crime, crime victimization, and witnessing crime using lenient correction for multiple comparisons

| | **Admin data** | | **Endline survey** | | |
|---|---|---|---|---|---|
| | **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 |
| | [-0.068, 0.074] | [0.011, 0.208] | [-0.077, 0.089] | [-0.098, 0.085] | [0.051, 0.256] |
| | (0.934) | (0.029) | (0.886) | (0.886) | (0.003) |
| Spillover | -0.038 | 0.083 | 0.026 | 0.013 | 0.186 |
| | [-0.097, 0.022] | [-0.003, 0.169] | [-0.034, 0.086] | [-0.061, 0.087] | [0.101, 0.270] |
| | (0.212) | (0.059) | (0.389) | (0.729) | (<0.001) |
| Individual controls | No | No | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes | Yes | Yes |
| 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 | | Yes | | | Yes |
| BH-spillover | | | | | Yes |
| Holm-treatment | | | | | Yes |
| Holm-spillover | | | | | Yes |

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. The dependent variable in columns 1 and 2 is an additive index of murders, armed robberies, thefts, illegal drug sales, and illegal possession of weapons. The dependent variable in columns 3 and 4 is a standardized additive index of victimization by vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, domestic violence, or extortion. The dependent variable in column 5 is a standardized additive index of witnessing vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, domestic violence, extortion, prostitution, illegal drug sales, illegal drug consumption, public alcohol consumption, or illegal possession of firearms in the past month. The frequency of crime witnessing was measured on a 1-4 Likert scale. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. Models 1 and 2 also include a lagged dependent variable and block-level controls for the average age, average years of education, and percentage of men on each block. Models 3-5 also include individual-level controls for age, gender, and years of education. ITT effect sizes in columns 1 and 2 are derived from WLS regressions as in equation 1. ITT effect sizes in columns 3-5 are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors in models 3-5 are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. We also report the Benjamini-Hochberg $q$-value for each model, and indicate whether each $p$-value falls below the corresponding $q$-value. We also indicate whether each $p$-value falls below the Holm-Bonferroni threshold. We assume that each model amounts to one hypothesis test.

Supplementary Table 7: Treatment effects on crime, crime victimization, and witnessing crime using stringent correction for multiple comparisons

| | **Admin data** | | **Endline survey** | | |
|---|---|---|---|---|---|
| | **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 |
| | [-0.068, 0.074] | [0.011, 0.208] | [-0.077, 0.089] | [-0.098, 0.085] | [0.051, 0.256] |
| | (0.934) | (0.029) | (0.886) | (0.886) | (0.003) |
| Spillover | -0.038 | 0.083 | 0.026 | 0.013 | 0.186 |
| | [-0.097, 0.022] | [-0.003, 0.169] | [-0.034, 0.086] | [-0.061, 0.087] | [0.101, 0.270] |
| | (0.212) | (0.059) | (0.389) | (0.729) | ($<$0.001) |
| Individual controls | No | No | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes | Yes | Yes |
| 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 | | Yes | | | Yes |
| BH-spillover | | | | | Yes |
| Holm-treatment | | | | | Yes |
| Holm-spillover | | | | | Yes |

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. The dependent variable in columns 1 and 2 is an additive index of murders, armed robberies, thefts, illegal drug sales, and illegal possession of weapons. The dependent variable in columns 3 and 4 is a standardized additive index of victimization by vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, domestic violence, or extortion. The dependent variable in column 5 is a standardized additive index of witnessing vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, domestic violence, extortion, prostitution, illegal drug sales, illegal drug consumption, public alcohol consumption, or illegal possession of firearms in the past month. The frequency of crime witnessing was measured on a 1-4 Likert scale. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. Models 1 and 2 also include a lagged dependent variable and block-level controls for the average age, average years of education, and percentage of men on each block. Models 3-5 also include individual-level controls for age, gender, and years of education. ITT effect sizes in columns 1 and 2 are derived from WLS regressions as in equation 1. ITT effect sizes in columns 3-5 are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors in models 3-5 are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. We also report the Benjamini-Hochberg $q$-value for each model, and indicate whether each $p$-value falls below the corresponding $q$-value. We also indicate whether each $p$-value falls below the Holm-Bonferroni threshold. We assume that each model amounts to two hypothesis tests.

Supplementary Table 8: Treatment effects on perceptions of security using lenient correction for multiple comparisons

| | Endline survey | | |
|---|---|---|---|
| | Perceptions of security | | |
| | **All safety** | **Personal safety** | **Business safety** |
| Treatment | -0.050 | -0.052 | 0.284 |
| | [-0.142, 0.043] | [-0.144, 0.041] | [0.056, 0.512] |
| | (0.292) | (0.271) | (0.015) |
| Spillover | -0.066 | -0.068 | 0.094 |
| | [-0.144, 0.013] | [-0.146, 0.010] | [-0.110, 0.297] |
| | (0.100) | (0.089) | (0.367) |
| Individual-level controls | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes |
| 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 | | | Yes |
| BH-spillover | | | |
| Holm-treatment | | | Yes |
| Holm-spillover | | | |

Notes: ITT on perceptions of safety for residents and business owners together (column 1), residents only (column 2), and business owners only (column 3) based on survey data. Residents and business owners were asked how safe they feel walking their blocks during the day and at night; how safe they feel talking on a smartphone on their block; and how worried they are about becoming victims of a violent or non-violent crime in the next two weeks. Perceptions of safety were measured on a 1-5 Likert scale. Residents and business owners were also asked whether they had taken precautions for fear of crime in the past month, including avoiding leaving their home or business at night; avoiding public transportation or recreation areas; prohibiting children from playing in the streets or attending school; considering moving to a different neighborhood; or changing jobs or schools. The dependent variable in models 1 and 2 is a standardized additive index based on responses to these questions. Business owners were also asked if they had changed their hours, hired private security guards, or closed their businesses for fear of crime. The dependent variable in model 3 is a standardized additive index based on responses to these questions. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. We also report the Benjamini-Hochberg $q$-value for each model, and indicate whether each $p$-value falls below the corresponding $q$-value. We also indicate whether each $p$-value falls below the Holm-Bonferroni threshold. We assume that each model amounts to one hypothesis test.

Supplementary Table 9: Treatment effects on perceptions of security using stringent correction for multiple comparisons

| | Endline survey | | |
|---|---|---|---|
| | Perceptions of security | | |
| | **All safety** | **Personal safety** | **Business safety** |
| Treatment | -0.050 | -0.052 | 0.284 |
| | [-0.142, 0.043] | [-0.144, 0.041] | [0.056, 0.512] |
| | (0.292) | (0.271) | (0.015) |
| Spillover | -0.066 | -0.068 | 0.094 |
| | [-0.144, 0.013] | [-0.146, 0.010] | [-0.110, 0.297] |
| | (0.100) | (0.089) | (0.367) |
| Individual-level controls | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes |
| 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 | | | Yes |
| BH-spillover | | | |
| Holm-treatment | | | Yes |
| Holm-spillover | | | |

Notes: ITT on perceptions of safety for residents and business owners together (column 1), residents only (column 2), and business owners only (column 3) based on survey data. Residents and business owners were asked how safe they feel walking their blocks during the day and at night; how safe they feel talking on a smartphone on their block; and how worried they are about becoming victims of a violent or non-violent crime in the next two weeks. Perceptions of safety were measured on a 1-5 Likert scale. Residents and business owners were also asked whether they had taken precautions for fear of crime in the past month, including avoiding leaving their home or business at night; avoiding public transportation or recreation areas; prohibiting children from playing in the streets or attending school; considering moving to a different neighborhood; or changing jobs or schools. The dependent variable in models 1 and 2 is a standardized additive index based on responses to these questions. Business owners were also asked if they had changed their hours, hired private security guards, or closed their businesses for fear of crime. The dependent variable in model 3 is a standardized additive index based on responses to these questions. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. We also report the Benjamini-Hochberg $q$-value for each model, and indicate whether each $p$-value falls below the corresponding $q$-value. We also indicate whether each $p$-value falls below the Holm-Bonferroni threshold. We assume that each model amounts to two hypothesis tests.

Supplementary Table 10: Treatment effects on abuses using lenient correction for multiple comparisons

| | Monitoring survey | | Endline survey | |
|---|---|---|---|---|
| | **Police abuse** | **Military abuse** | **Police abuse** | **Military abuse** |
| Treatment | 0.037 | 0.010 | 0.011 | -0.001 |
| | [0.016, 0.058] | [-0.001, 0.022] | [-0.015, 0.037] | [-0.009, 0.006] |
| | (0.001) | (0.071) | (0.417) | (0.716) |
| Spillover | 0.016 | 0.001 | 0.030 | 0.002 |
| | [-0.005, 0.037] | [-0.005, 0.007] | [0.007, 0.053] | [-0.005, 0.009] |
| | (0.137) | (0.778) | (0.011) | (0.544) |
| Individual-level controls | Yes | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes | Yes |
| Observations | 1970 | 1970 | 7908 | 7908 |
| $R^2$ | 0.05 | 0.05 | 0.05 | 0.01 |
| Control mean | 0.015 | 0.000 | 0.114 | 0.012 |
| qval-treatment | 0.003 | 0.142 | 0.556 | 0.716 |
| qval-spillover | 0.275 | 0.778 | 0.044 | 0.726 |
| BH-treatment | Yes | | | |
| BH-spillover | | | Yes | |
| Holm-treatment | Yes | | | |
| Holm-spillover | | | Yes | |

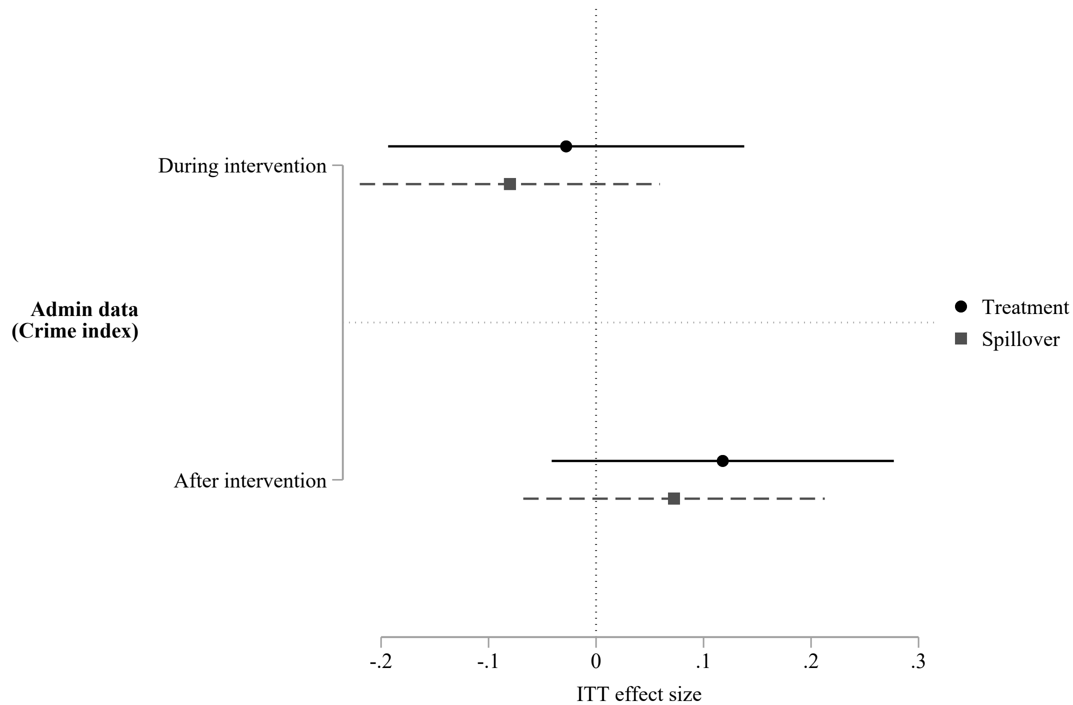
Notes: ITT on abuses committed by police officers and soldiers based on monitoring (columns 1 and 2) and survey data (columns 3 and 4). The dependent variable in all models is an indicator for any incidents of verbal or physical abuse on the block. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age and gender; models in columns 3 and 4 include an additional individual-level control for years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. We also report the Benjamini-Hochberg $q$-value for each model, and indicate whether each $p$-value falls below the corresponding $q$-value. We also indicate whether each $p$-value falls below the Holm-Bonferroni threshold. We assume that each model amounts to one hypothesis test.

Supplementary Table 11: Treatment effects on abuses using stringent correction for multiple comparisons

| | Monitoring survey | | Endline survey | |
|---|---|---|---|---|
| | **Police abuse** | **Military abuse** | **Police abuse** | **Military abuse** |
| Treatment | 0.037 | 0.010 | 0.011 | -0.001 |
| | [0.016, 0.058] | [-0.001, 0.022] | [-0.015, 0.037] | [-0.009, 0.006] |
| | (0.001) | (0.071) | (0.417) | (0.716) |
| Spillover | 0.016 | 0.001 | 0.030 | 0.002 |
| | [-0.005, 0.037] | [-0.005, 0.007] | [0.007, 0.053] | [-0.005, 0.009] |
| | (0.137) | (0.778) | (0.011) | (0.544) |
| Individual-level controls | Yes | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes | Yes |
| Observations | 1970 | 1970 | 7908 | 7908 |
| $R^2$ | 0.05 | 0.05 | 0.05 | 0.01 |
| Control mean | 0.015 | 0.000 | 0.114 | 0.012 |
| qval-treatment | 0.005 | 0.189 | 0.668 | 0.778 |
| qval-spillover | 0.275 | 0.778 | 0.044 | 0.726 |
| BH-treatment | Yes | | | |
| BH-spillover | | | Yes | |
| Holm-treatment | Yes | | | |
| Holm-spillover | | | Yes | |

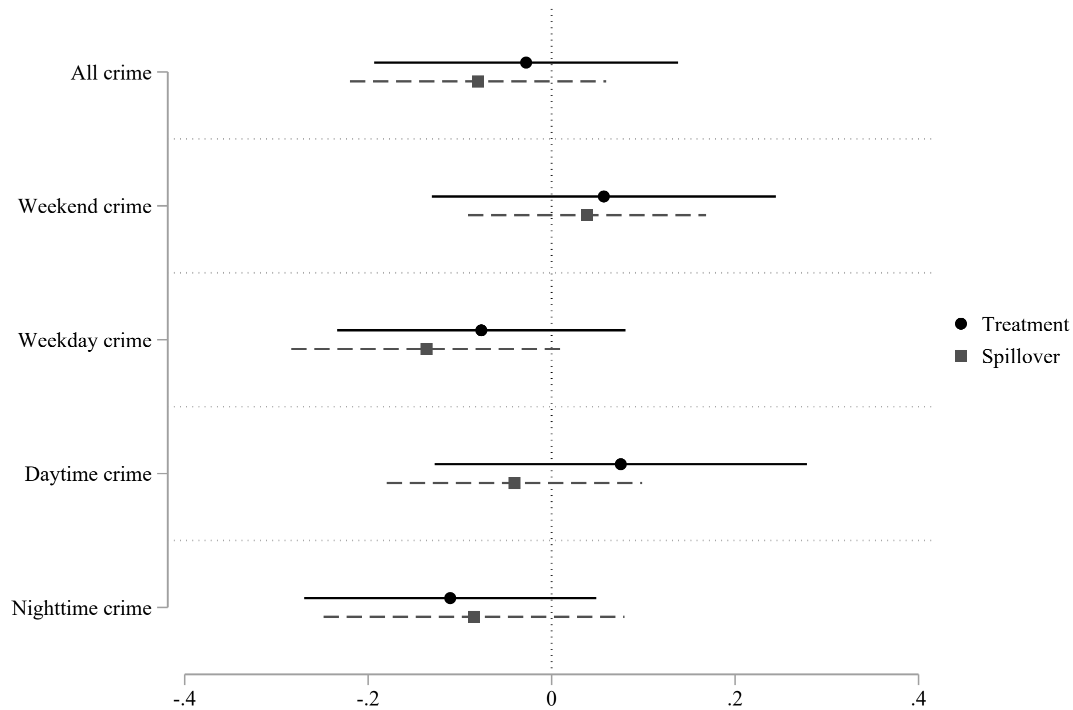
Notes: ITT on abuses committed by police officers and soldiers based on monitoring (columns 1 and 2) and survey data (columns 3 and 4). The dependent variable in all models is an indicator for any incidents of verbal or physical abuse on the block. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age and gender; models in columns 3 and 4 include an additional individual-level control for years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. We also report the Benjamini-Hochberg $q$-value for each model, and indicate whether each $p$-value falls below the corresponding $q$-value. We also indicate whether each $p$-value falls below the Holm-Bonferroni threshold. We assume that each model amounts to two hypothesis tests.

Supplementary Figure 6: Treatment effects on crime using weighted standardized index



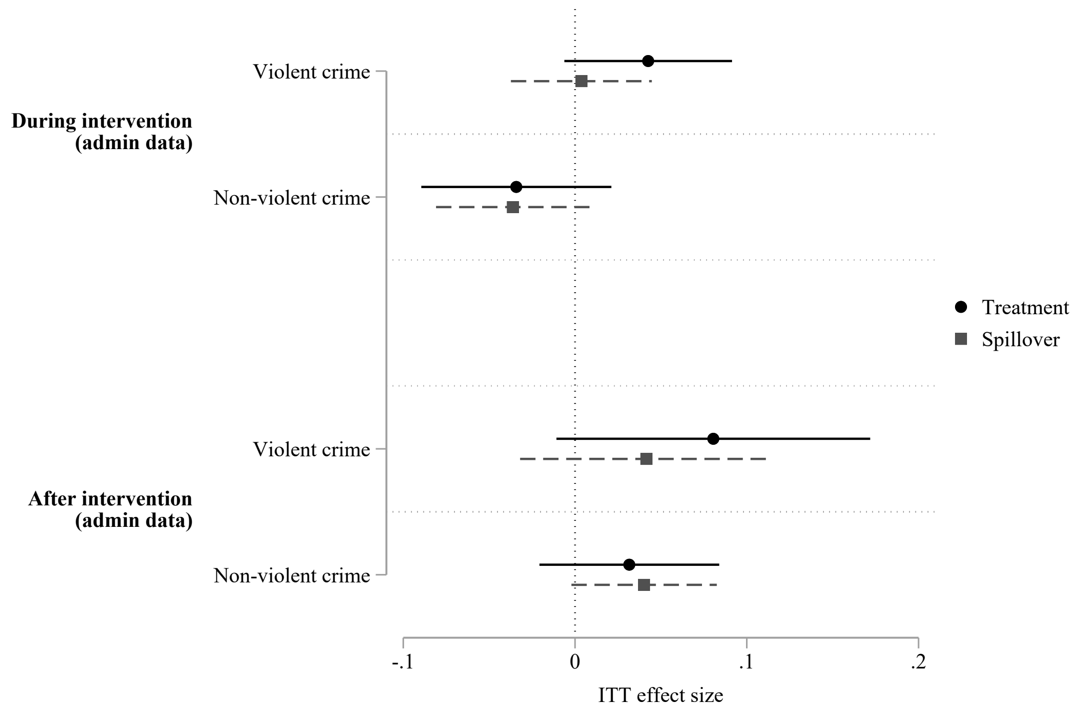

Notes: ITT on crime during (top panel) and after (bottom panel) the intervention based on administrative data. The dependent variable in all models is an additive index of murders, armed robberies, thefts, illegal drug sales, and illegal possession of weapons, weighted by the average prison sentence associated with each crime under Colombian law. All models include neighborhood fixed effects, a lagged dependent variable, and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include block-level controls for the average age, average years of education, and percentage of men on each block. ITT effect sizes are derived from WLS regressions as in equation 1. Observations are weighted by the inverse probability of assignment to their realized treatment status. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For crime "during intervention" ($N = 1,167$): $\beta = -0.028$, CI = [-0.193, 0.138], $p = 0.742$, RI $p = 0.822$ on treatment blocks; $\beta = -0.080$, CI = [-0.220, 0.059], $p = 0.260$, RI $p = 0.413$ on spillover blocks. For crime "after intervention" ($N = 1,167$): $\beta = 0.118$, CI = [-0.041, 0.277], $p = 0.146$, RI $p = 0.345$ on treatment blocks; $\beta = 0.073$, CI = [-0.068, 0.213], $p = 0.311$, RI $p = 0.439$ on spillover blocks.

Supplementary Figure 7: Treatment effects on crime by day and time using weighted standardized index




Notes: ITT on crime during the intervention based on administrative data, disaggregated by day and time. The dependent variable is an additive index of murders, armed robberies, thefts, illegal drug sales, and illegal possession of weapons, weighted by the average prison sentence associated with each crime under Colombian law. All models include neighborhood fixed effects, a lagged dependent variable, and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include block-level controls for the average age, average years of education, and percentage of men on each block. ITT effect sizes are derived from WLS regressions as in equation 1. Observations are weighted by the inverse probability of assignment to their realized treatment status. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "all crime" ($N = 1,167$): $\beta = -0.028$, CI = [-0.193, 0.138], $p = 0.742$, RI $p = 0.822$ on treatment blocks; $\beta = -0.080$, CI = [-0.220, 0.059], $p = 0.260$, RI $p = 0.413$ on spillover blocks. For "weekend crime" ($N = 1,167$): $\beta = 0.057$, CI = [-0.131, 0.244], $p = 0.551$, RI $p = 0.647$ on treatment blocks; $\beta = 0.039$, CI = [-0.091, 0.168], $p = 0.559$, RI $p = 0.724$ on spillover blocks. For "weekday crime" ($N = 1,167$): $\beta = -0.077$, CI = [-0.234, 0.081], $p = 0.339$, RI $p = 0.523$ on treatment blocks; $\beta = -0.136$, CI = [-0.284, 0.012], $p = 0.071$, RI $p = 0.177$ on spillover blocks. For "daytime crime" ($N = 1,167$): $\beta = 0.075$, CI = [-0.127, 0.278], $p = 0.466$, RI $p = 0.617$ on treatment blocks; $\beta = -0.041$, CI = [-0.180, 0.099], $p = 0.568$, RI $p = 0.737$ on spillover blocks. For "nighttime crime" ($N = 1,167$): $\beta = -0.111$, CI = [-0.270, 0.049], $p = 0.173$, RI $p = 0.231$ on treatment blocks; $\beta = -0.085$, CI = [-0.249, 0.079], $p = 0.312$, RI $p = 0.256$ on spillover blocks.

Supplementary Figure 8: Treatment effects on violent and non-violent crime in admin data




Notes: ITT on crime during (top panel) and after (bottom panel) the intervention based on administrative data. The dependent variable in the top panel is an additive index of murders and armed robberies. The dependent variable in the bottom panel is an additive index of thefts, illegal drug sales, and illegal possession of weapons. All models include neighborhood fixed effects, a lagged dependent variable, and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include block-level controls for the average age, average years of education, and percentage of men on each block. ITT effect sizes are derived from WLS regressions as in equation 1. Observations are weighted by the inverse probability of assignment to their realized treatment status. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "violent crime during intervention" ($N = 1,167$): $\beta = 0.043$, CI = [-0.006, 0.091], $p = 0.087$, RI $p = 0.276$ on treatment blocks; $\beta = 0.004$, CI = [-0.037, 0.045], $p = 0.861$, RI $p = 0.904$ on spillover blocks. For "non-violent crime during intervention" ($N = 1,167$): $\beta = -0.034$, CI = [-0.090, 0.021], $p = 0.226$, RI $p = 0.419$ on treatment blocks; $\beta = -0.036$, CI = [-0.081, 0.008], $p = 0.112$, RI $p = 0.293$ on spillover blocks. For "violent crime after intervention" ($N = 1,167$): $\beta = 0.081$, CI = [-0.011, 0.172], $p = 0.084$, RI $p = 0.208$ on treatment blocks; $\beta = 0.042$, CI = [-0.032, 0.115], $p = 0.268$, RI $p = 0.386$ on spillover blocks. For "non-violent crime after intervention" ($N = 1,167$): $\beta = 0.032$, CI = [-0.021, 0.084], $p = 0.236$, RI $p = 0.449$ on treatment blocks; $\beta = 0.040$, CI = [-0.002, 0.083], $p = 0.063$, RI $p = 0.219$ on spillover blocks.

Supplementary Table 12: Treatment effects on violent crime victimization and witnessing in survey data

| | **Endline survey** | | |
|---|---|---|---|
| | **Victim** | | **Witness** |
| | During intervention | After intervention | After intervention |
| Treatment | -0.016 | -0.008 | 0.146 |
| | [-0.110, 0.077] | [-0.097, 0.081] | [0.048, 0.245] |
| | (0.731) | (0.862) | (0.004) |
| Spillover | -0.011 | 0.019 | 0.172 |
| | [-0.085, 0.062] | [-0.052, 0.090] | [0.092, 0.252] |
| | (0.761) | (0.598) | (<0.001) |
| Individual controls | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes |
| Block-level controls | No | No | No |
| Observations | 7883 | 7883 | 7898 |
| $R^2$ | 0.01 | 0.02 | 0.10 |
| Control mean | -0.005 | -0.009 | -0.102 |
| RI p-value (treatment) | 0.794 | 0.895 | 0.046 |
| RI p-value (spillover) | 0.829 | 0.702 | 0.004 |

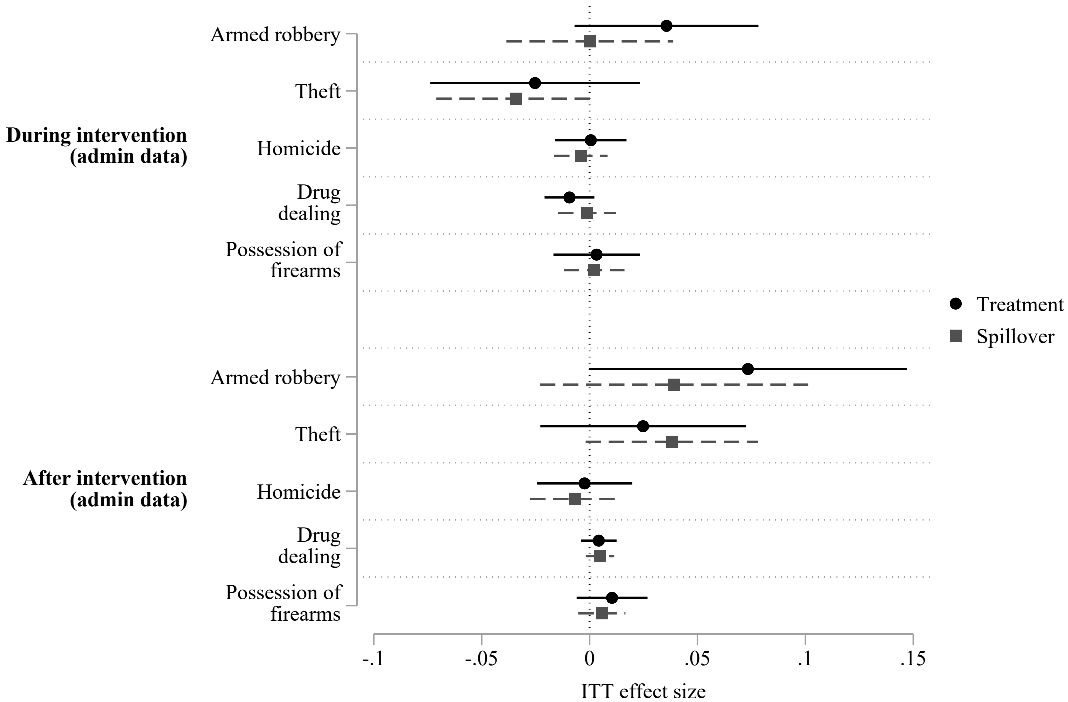
Notes: ITT on violent crime victimization during (column 1) and after (column 2) the intervention based on survey data, and violent crime witnessing after the intervention (column 3) based on survey data. The dependent variable in columns 1 and 2 is a standardized additive index of victimization by armed robbery, homicide, or attempted homicide. The dependent variable in column 3 is a standardized additive index of witnessing armed robbery, homicide, or attempted homicide in the past month. The frequency of crime witnessing was measured on a 1-4 Likert scale. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Table 13: Treatment effects on non-violent crime victimization and witnessing in survey data

| | Endline survey | | |
|---|---|---|---|
| | Victim | | Witness |
| | During intervention | After intervention | After intervention |
| Treatment | 0.014 | -0.008 | 0.148 |
| | [-0.061, 0.089] | [-0.094, 0.077] | [0.045, 0.250] |
| | (0.713) | (0.850) | (0.005) |
| Spillover | 0.041 | 0.006 | 0.180 |
| | [-0.018, 0.100] | [-0.065, 0.077] | [0.095, 0.265] |
| | (0.169) | (0.874) | ($<$0.001) |
| Individual controls | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes |
| Block-level controls | No | No | No |
| Observations | 7865 | 7865 | 7840 |
| $R^2$ | 0.03 | 0.03 | 0.11 |
| Control mean | -0.024 | -0.015 | -0.118 |
| RI p-value (treatment) | 0.811 | 0.889 | 0.041 |
| RI p-value (spillover) | 0.360 | 0.903 | 0.002 |

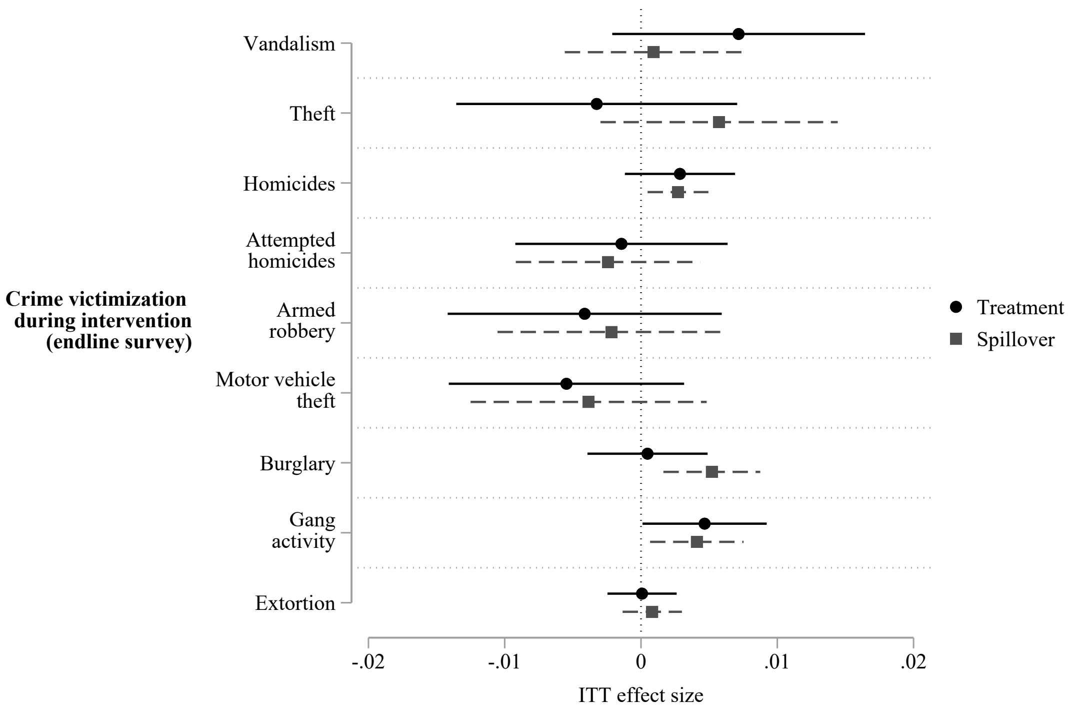
Notes: ITT on non-violent crime victimization during (column 1) and after (column 2) the intervention based on survey data, and non-violent crime witnessing after the intervention (column 3) based on survey data. The dependent variable in columns 1 and 2 is a standardized additive index of victimization by vandalism, burglary, theft, motor vehicle theft, gang activity, or extortion. The dependent variable in column 3 is a standardized additive index of witnessing vandalism, burglary, theft, motor vehicle theft, gang activity, domestic violence, extortion, prostitution, illegal drug sales, illegal drug consumption, public alcohol consumption, or illegal possession of a firearm in the past month. The frequency of crime witnessing was measured on a 1-4 Likert scale. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Figure 9: Treatment effects on crime in administrative data disaggregated by type



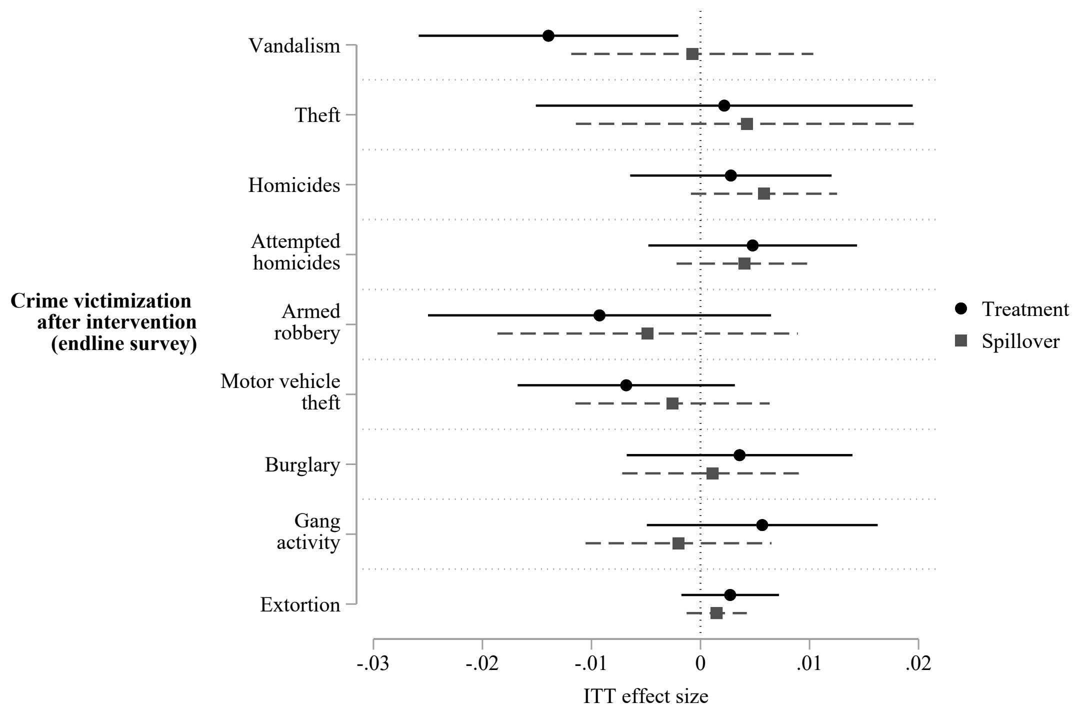

Notes: ITT on crime during (top panel) and after (bottom panel) the intervention based on administrative data, disaggregated by type. The dependent variables are additive indices of murders, armed robberies, thefts, illegal drug sales, and illegal possession of weapons. All models include neighborhood fixed effects, a lagged dependent variable, and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include block-level controls for the average age, average years of education, and percentage of men on each block. ITT effect sizes are derived from WLS regressions as in equation 1. Observations are weighted by the inverse probability of assignment to their realized treatment status. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "armed robbery during intervention" ($N = 1,167$): $\beta = 0.036$, CI = [-0.007, 0.078], $p = 0.101$, RI $p = 0.319$ on treatment blocks; $\beta = 0.000$, CI = [-0.039, 0.039], $p = 0.995$, RI $p = 0.997$ on spillover blocks. For "theft during intervention" ($N = 1,167$): $\beta = -0.025$, CI = [-0.074, 0.023], $p = 0.307$, RI $p = 0.517$ on treatment blocks; $\beta = -0.034$, CI = [-0.071, 0.003], $p = 0.071$, RI $p = 0.290$ on spillover blocks. For "homicide during intervention" ($N = 1,167$): $\beta = 0.001$, CI = [-0.016, 0.017], $p = 0.945$, RI $p = 0.960$ on treatment blocks; $\beta = -0.004$, CI = [-0.016, 0.008], $p = 0.520$, RI $p = 0.700$ on spillover blocks. For "drug dealing during intervention" ($N = 1,167$): $\beta = -0.009$, CI = [-0.021, 0.002], $p = 0.114$, RI $p = 0.122$ on treatment blocks; $\beta = -0.001$, CI = [-0.015, 0.012], $p = 0.860$, RI $p = 0.787$ on spillover blocks. For "possession of firearms during intervention" ($N = 1,167$): $\beta = 0.003$, CI = [-0.017, 0.023], $p = 0.750$, RI $p = 0.702$ on treatment blocks; $\beta = 0.002$, CI = [-0.012, 0.016], $p = 0.765$, RI $p = 0.844$ on spillover blocks. For "armed robbery after intervention" ($N = 1,167$): $\beta = 0.073$, CI = [-0.000, 0.147], $p = 0.051$, RI $p = 0.192$ on treatment blocks; $\beta = 0.039$, CI = [-0.023, 0.101], $p = 0.217$, RI $p = 0.358$ on spillover blocks. For "theft after intervention" ($N = 1,167$): $\beta = 0.025$, CI = [-0.023, 0.072], $p = 0.307$, RI $p = 0.486$ on treatment blocks; $\beta = 0.038$, CI = [-0.002, 0.078], $p = 0.062$, RI $p = 0.179$ on spillover blocks. For "homicide after intervention" ($N = 1,167$): $\beta = -0.002$, CI = [-0.024, 0.020], $p = 0.838$, RI $p = 0.895$ on treatment blocks; $\beta = -0.007$, CI = [-0.028, 0.014], $p = 0.514$, RI $p = 0.657$ on spillover blocks. For "drug dealing after intervention" ($N = 1,167$): $\beta = 0.004$, CI = [-0.004, 0.013], $p = 0.312$, RI $p = 0.562$ on treatment blocks; $\beta = 0.005$, CI = [-0.002, 0.011], $p = 0.153$, RI $p = 0.322$ on spillover blocks. For "weapons possession after intervention" ($N = 1,167$): $\beta = 0.010$, CI = [-0.006, 0.027], $p = 0.213$, RI $p = 0.459$ on treatment blocks; $\beta = 0.006$, CI = [-0.005, 0.017], $p = 0.307$, RI $p = 0.654$ on spillover blocks.

Supplementary Figure 10: Treatment effects on crime victimization in survey data disaggregated by type



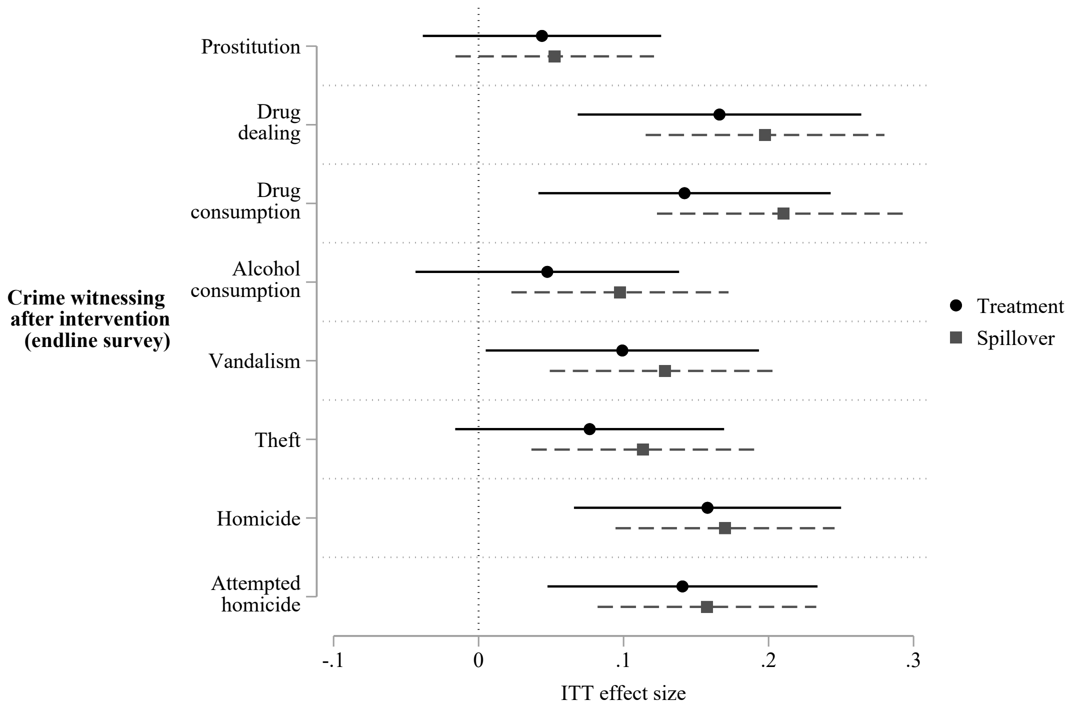

Notes: ITT on crime victimization during the intervention based on survey data, disaggregated by type. The dependent variables are indicators for victimization by vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, or extortion. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "vandalism" ($N = 7,896$): $\beta = 0.007$, CI = [-0.002, 0.016], $p = 0.130$, RI $p = 0.314$ on treatment blocks; $\beta = 0.001$, CI = [-0.006, 0.007], $p = 0.779$, RI $p = 0.866$ on spillover blocks. For "thefts" ($N = 7,893$): $\beta = -0.003$, CI = [-0.014, 0.007], $p = 0.536$, RI $p = 0.681$ on treatment blocks; $\beta = 0.006$, CI = [-0.003, 0.014], $p = 0.197$, RI $p = 0.365$ on spillover blocks. For "homicides" ($N = 7,903$): $\beta = 0.003$, CI = [-0.001, 0.007], $p = 0.165$, RI $p = 0.342$ on treatment blocks; $\beta = 0.003$, CI = [0.000, 0.005], $p = 0.017$, RI $p = 0.268$ on spillover blocks. For "attempted homicides" ($N = 7,901$): $\beta = -0.001$, CI = [-0.009, 0.006], $p = 0.718$, RI $p = 0.746$ on treatment blocks; $\beta = -0.002$, CI = [-0.009, 0.004], $p = 0.482$, RI $p = 0.547$ on spillover blocks. For "armed robbery" ($N = 7,895$): $\beta = -0.004$, CI = [-0.014, 0.006], $p = 0.420$, RI $p = 0.582$ on treatment blocks; $\beta = -0.002$, CI = [-0.011, 0.006], $p = 0.611$, RI $p = 0.715$ on spillover blocks. For "motor vehicle theft" ($N = 7,899$): $\beta = -0.005$, CI = [-0.014, 0.003], $p = 0.214$, RI $p = 0.308$ on treatment blocks; $\beta = -0.004$, CI = [-0.013, 0.005], $p = 0.384$, RI $p = 0.336$ on spillover blocks. For "burglary" ($N = 7,902$): $\beta = 0.000$, CI = [-0.004, 0.005], $p = 0.831$, RI $p = 0.888$ on treatment blocks; $\beta = 0.005$, CI = [0.002, 0.009], $p = 0.004$, RI $p = 0.046$ on spillover blocks. For "gang activity" ($N = 7,903$): $\beta = 0.005$, CI = [0.000, 0.009], $p = 0.045$, RI $p = 0.227$ on treatment blocks; $\beta = 0.004$, CI = [0.001, 0.008], $p = 0.019$, RI $p = 0.172$ on spillover blocks. For "extortion" ($N = 7,906$): $\beta = 0.000$, CI = [-0.002, 0.003], $p = 0.951$, RI $p = 0.968$ on treatment blocks; $\beta = 0.001$, CI = [-0.001, 0.003], $p = 0.457$, RI $p = 0.637$ on spillover blocks.

Supplementary Figure 10: Treatment effects on crime victimization in survey data disaggregated by type (cont.)



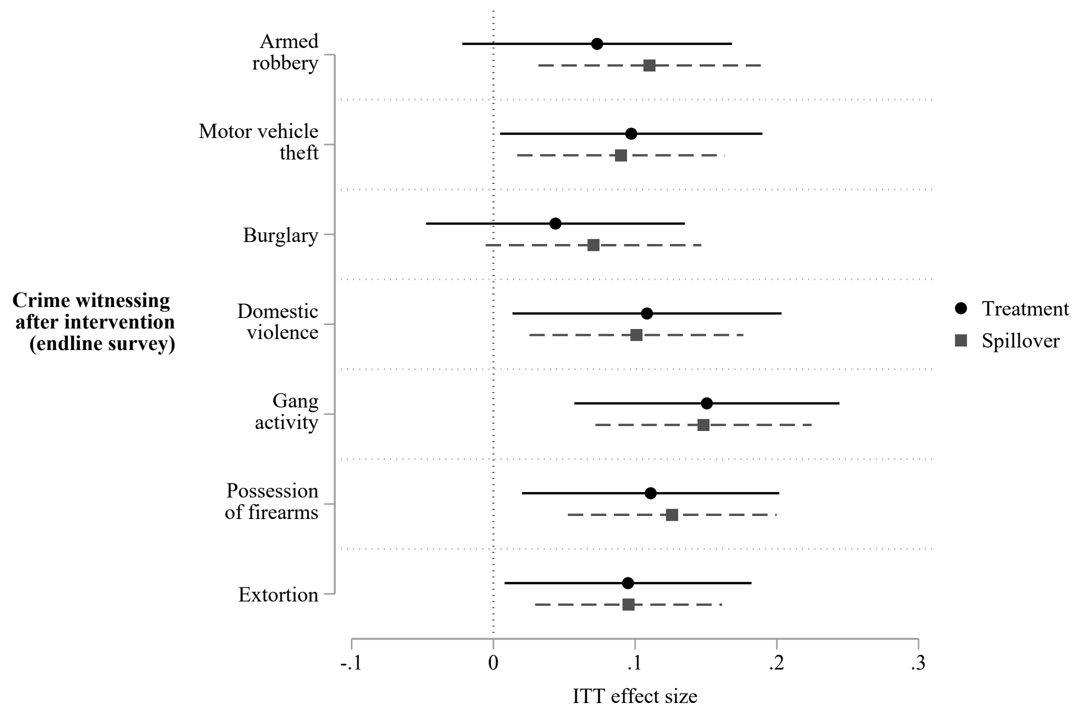

Notes: ITT on crime victimization after the intervention based on survey data, disaggregated by type. The dependent variables are indicators for victimization by vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, or extortion. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "vandalism" ($N = 7,896$): $\beta = -0.014$, CI = [-0.026, -0.002], $p = 0.022$, RI $p = 0.070$ on treatment blocks; $\beta = -0.001$, CI = [-0.012, 0.010], $p = 0.894$, RI $p = 0.903$ on spillover blocks. For "thefts" ($N = 7,893$): $\beta = 0.002$, CI = [-0.015, 0.019], $p = 0.804$, RI $p = 0.850$ on treatment blocks; $\beta = 0.004$, CI = [-0.011, 0.020], $p = 0.594$, RI $p = 0.635$ on spillover blocks. For "homicides" ($N = 7,903$): $\beta = 0.003$, CI = [-0.006, 0.012], $p = 0.554$, RI $p = 0.677$ on treatment blocks; $\beta = 0.006$, CI = [-0.001, 0.013], $p = 0.089$, RI $p = 0.246$ on spillover blocks. For "attempted homicides" ($N = 7,901$): $\beta = 0.005$, CI = [-0.005, 0.014], $p = 0.327$, RI $p = 0.467$ on treatment blocks; $\beta = 0.004$, CI = [-0.002, 0.010], $p = 0.205$, RI $p = 0.471$ on spillover blocks. For "armed robbery" ($N = 7,895$): $\beta = -0.009$, CI = [-0.025, 0.006], $p = 0.248$, RI $p = 0.353$ on treatment blocks; $\beta = -0.005$, CI = [-0.019, 0.009], $p = 0.490$, RI $p = 0.551$ on spillover blocks. For "motor vehicle theft" ($N = 7,899$): $\beta = -0.007$, CI = [-0.017, 0.003], $p = 0.180$, RI $p = 0.303$ on treatment blocks; $\beta = -0.003$, CI = [-0.011, 0.006], $p = 0.572$, RI $p = 0.599$ on spillover blocks. For "burglary" ($N = 7,902$): $\beta = 0.004$, CI = [-0.007, 0.014], $p = 0.497$, RI $p = 0.590$ on treatment blocks; $\beta = 0.001$, CI = [-0.007, 0.009], $p = 0.791$, RI $p = 0.824$ on spillover blocks. For "gang activity" ($N = 7,903$): $\beta = 0.006$, CI = [-0.005, 0.016], $p = 0.294$, RI $p = 0.472$ on treatment blocks; $\beta = -0.002$, CI = [-0.011, 0.007], $p = 0.644$, RI $p = 0.741$ on spillover blocks. For "extortion" ($N = 7,906$): $\beta = 0.003$, CI = [-0.002, 0.007], $p = 0.233$, RI $p = 0.441$ on treatment blocks; $\beta = 0.001$, CI = [-0.001, 0.004], $p = 0.292$, RI $p = 0.625$ on spillover blocks.

Supplementary Figure 11: Treatment effects on crime witnessing in survey data disaggregated by type



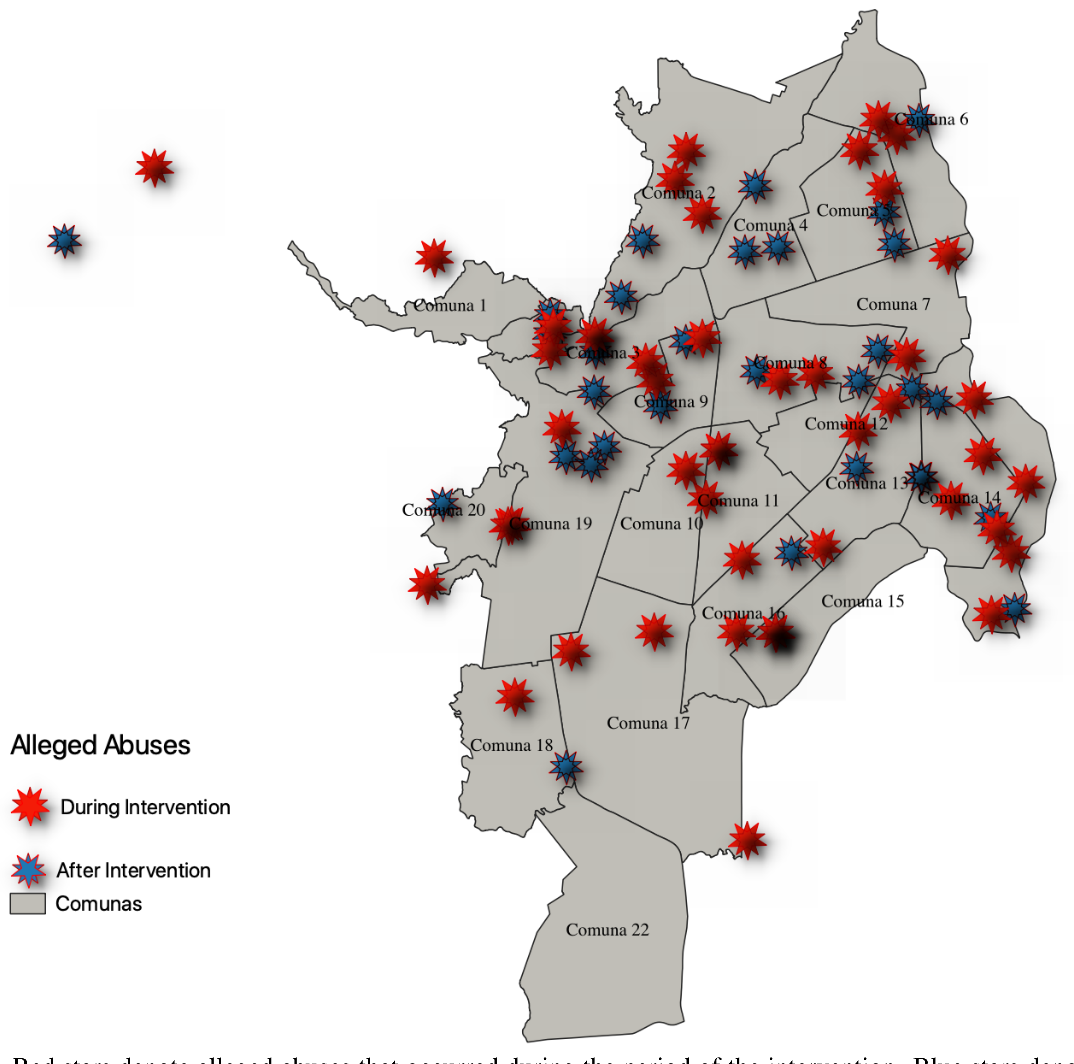

Notes: ITT on crime witnessing after the intervention based on survey data, disaggregated by type. The dependent variables are standardized indices of witnessing vandalism, theft, homicide, attempted homicide, prostitution, illegal drug sales, illegal drug consumption, or public alcohol consumption in the past month. The frequency of crime witnessing was measured on a 1-4 Likert scale. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "prostitution" ($N = 7,903$): $\beta = 0.044$, CI = [-0.038, 0.126], $p = 0.297$, RI $p = 0.487$ on treatment blocks; $\beta = 0.053$, CI = [-0.016, 0.121], $p = 0.133$, RI $p = 0.303$ on spillover blocks. For "drug dealing" ($N = 7,880$): $\beta = 0.166$, CI = [0.068, 0.264], $p = 0.001$, RI $p = 0.026$ on treatment blocks; $\beta = 0.198$, CI = [0.115, 0.280], $p = 0.000$, RI $p = 0.001$ on spillover blocks. For "drug consumption" ($N = 7,899$): $\beta = 0.142$, CI = [0.041, 0.243], $p = 0.006$, RI $p = 0.048$ on treatment blocks; $\beta = 0.210$, CI = [0.123, 0.298], $p = 0.000$, RI $p = 0.000$ on spillover blocks. For "alcohol consumption" ($N = 7,900$): $\beta = 0.047$, CI = [-0.044, 0.138], $p = 0.307$, RI $p = 0.460$ on treatment blocks; $\beta = 0.098$, CI = [0.023, 0.172], $p = 0.011$, RI $p = 0.058$ on spillover blocks. For "vandalism" ($N = 7,903$): $\beta = 0.099$, CI = [0.005, 0.193], $p = 0.039$, RI $p = 0.121$ on treatment blocks; $\beta = 0.129$, CI = [0.049, 0.208], $p = 0.002$, RI $p = 0.012$ on spillover blocks. For "theft" ($N = 7,899$): $\beta = 0.077$, CI = [-0.016, 0.169], $p = 0.105$, RI $p = 0.261$ on treatment blocks; $\beta = 0.113$, CI = [0.036, 0.190], $p = 0.004$, RI $p = 0.034$ on spillover blocks. For "homicide" ($N = 7,901$): $\beta = 0.158$, CI = [0.066, 0.250], $p = 0.001$, RI $p = 0.027$ on treatment blocks; $\beta = 0.170$, CI = [0.094, 0.246], $p = 0.000$, RI $p = 0.003$ on spillover blocks. For "attempted homicide" ($N = 7,900$): $\beta = 0.141$, CI = [0.047, 0.234], $p = 0.003$, RI $p = 0.041$ on treatment blocks; $\beta = 0.157$, CI = [0.082, 0.233], $p = 0.000$, RI $p = 0.004$ on spillover blocks.

Supplementary Figure 11: Treatment effects on crime witnessing in survey data disaggregated by type (cont.)



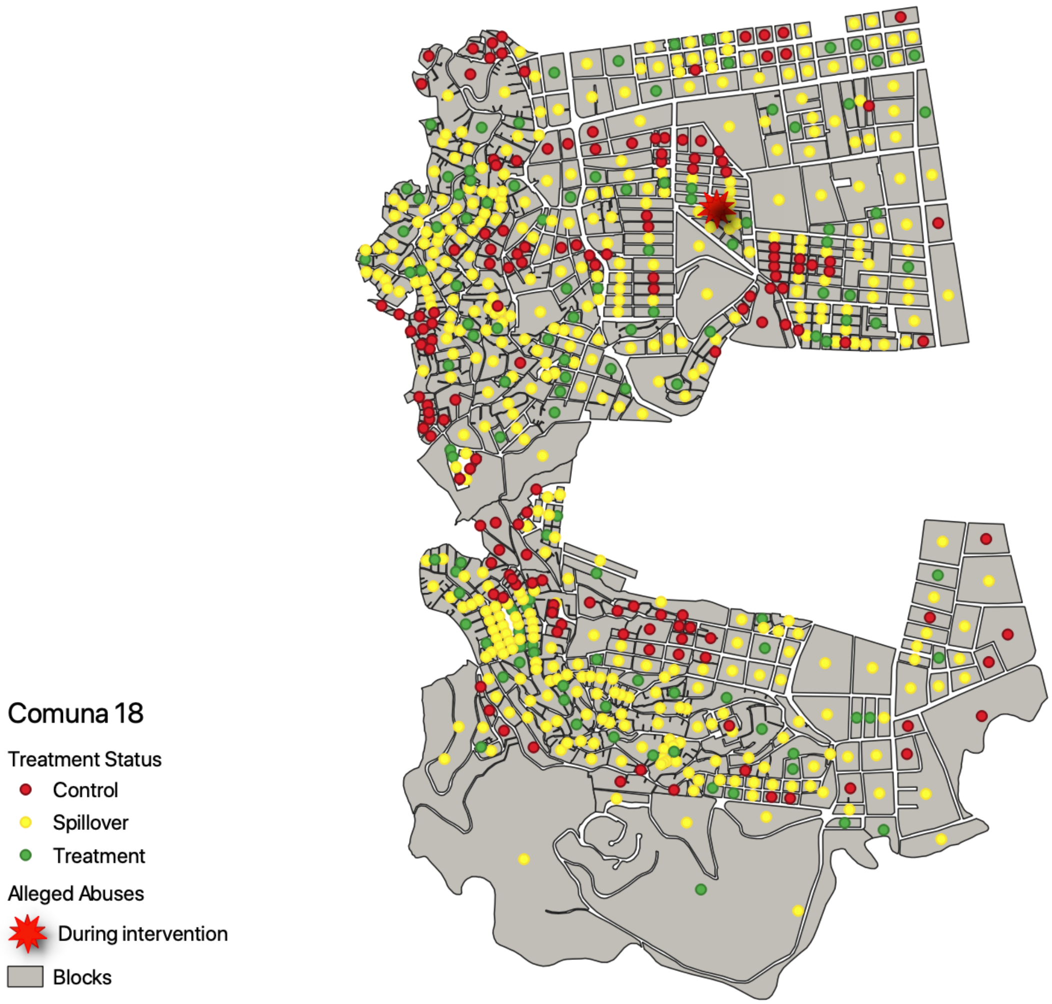

Notes: ITT on crime witnessing after the intervention based on survey data, disaggregated by type. The dependent variables are standardized indices of witnessing armed robbery, burglary, motor vehicle theft, gang activity, domestic violence, extortion, or illegal possession of a firearm in the past month. The frequency of crime witnessing was measured on a 1-4 Likert scale. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "armed robbery" ($N = 7,902$): $\beta = 0.073$, CI = [-0.022, 0.168], $p = 0.131$, RI $p = 0.314$ on treatment blocks; $\beta = 0.110$, CI = [0.032, 0.189], $p = 0.006$, RI $p = 0.055$ on spillover blocks. For "motor vehicle theft" ($N = 7,900$): $\beta = 0.097$, CI = [0.005, 0.190], $p = 0.039$, RI $p = 0.173$ on treatment blocks; $\beta = 0.090$, CI = [0.017, 0.163], $p = 0.016$, RI $p = 0.112$ on spillover blocks. For "burglary" ($N = 7,900$): $\beta = 0.044$, CI = [-0.047, 0.135], $p = 0.347$, RI $p = 0.507$ on treatment blocks; $\beta = 0.071$, CI = [-0.005, 0.147], $p = 0.069$, RI $p = 0.171$ on spillover blocks. For "domestic violence" ($N = 7,904$): $\beta = 0.108$, CI = [0.013, 0.203], $p = 0.025$, RI $p = 0.107$ on treatment blocks; $\beta = 0.101$, CI = [0.025, 0.176], $p = 0.009$, RI $p = 0.057$ on spillover blocks. For "gang activity" ($N = 7,887$): $\beta = 0.151$, CI = [0.057, 0.244], $p = 0.002$, RI $p = 0.028$ on treatment blocks; $\beta = 0.148$, CI = [0.072, 0.225], $p = 0.000$, RI $p = 0.005$ on spillover blocks. For "possession of firearms" ($N = 7,889$): $\beta = 0.111$, CI = [0.020, 0.202], $p = 0.017$, RI $p = 0.093$ on treatment blocks; $\beta = 0.126$, CI = [0.053, 0.200], $p = 0.001$, RI $p = 0.015$ on spillover blocks. For "extortion" ($N = 7,891$): $\beta = 0.095$, CI = [0.008, 0.182], $p = 0.033$, RI $p = 0.165$ on treatment blocks; $\beta = 0.095$, CI = [0.029, 0.161], $p = 0.005$, RI $p = 0.074$ on spillover blocks.

Supplementary Figure 12: Alleged abuses by state security forces in Cali as reported to Attorney-General's Office, September 30, 2019–December 31, 2019



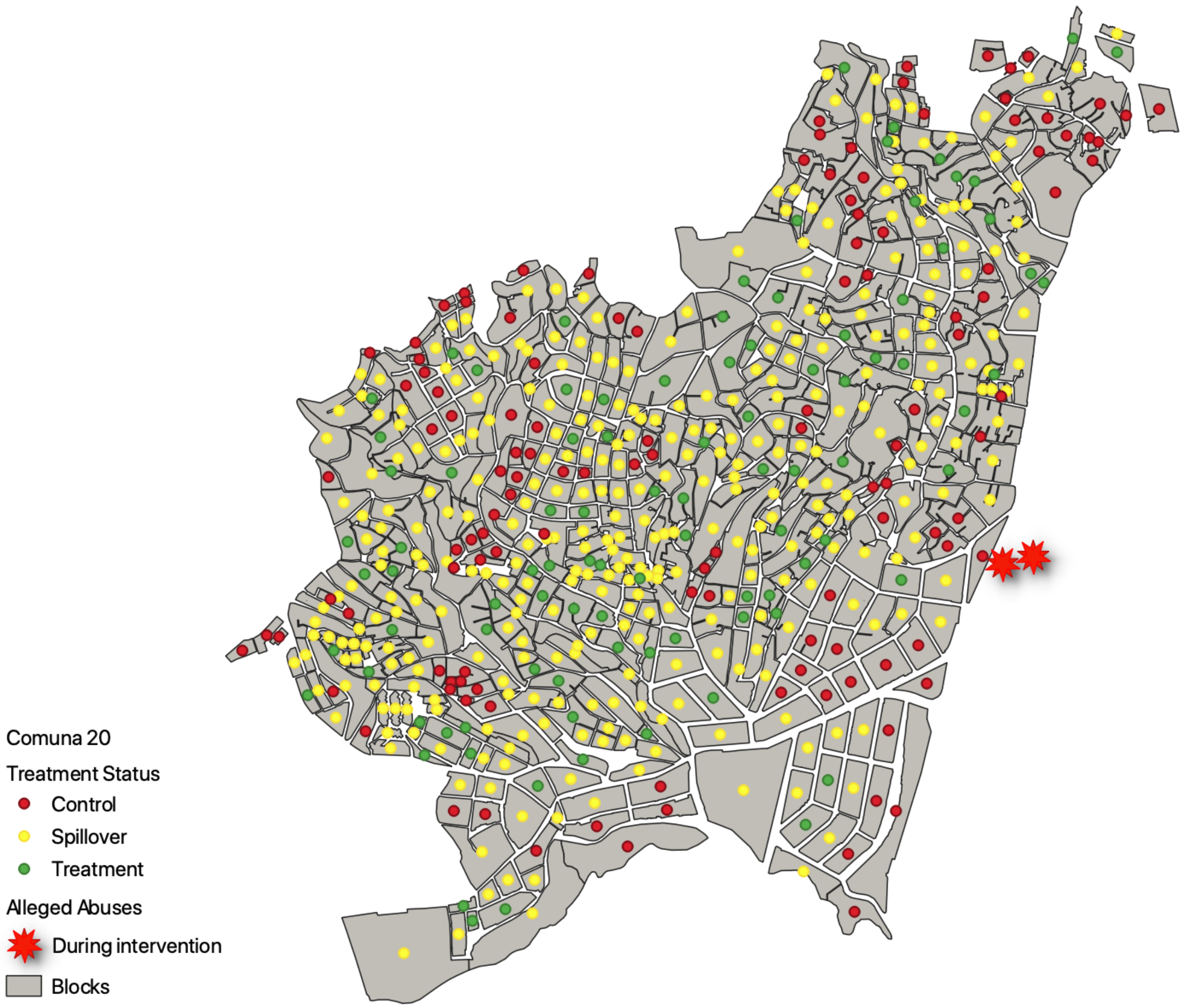

Notes: Red stars denote alleged abuses that occurred during the period of the intervention. Blue stars denote alleged abuses that occurred after the period of the intervention. Source: Instituto Geográfico Agustín Codazzi - Igac (underlying map) and Attorney-General's Office (abuse data).

Supplementary Figure 13: Alleged abuses by state security forces in commune 18 as reported to Attorney-General's Office, September 30–November 18, 2019



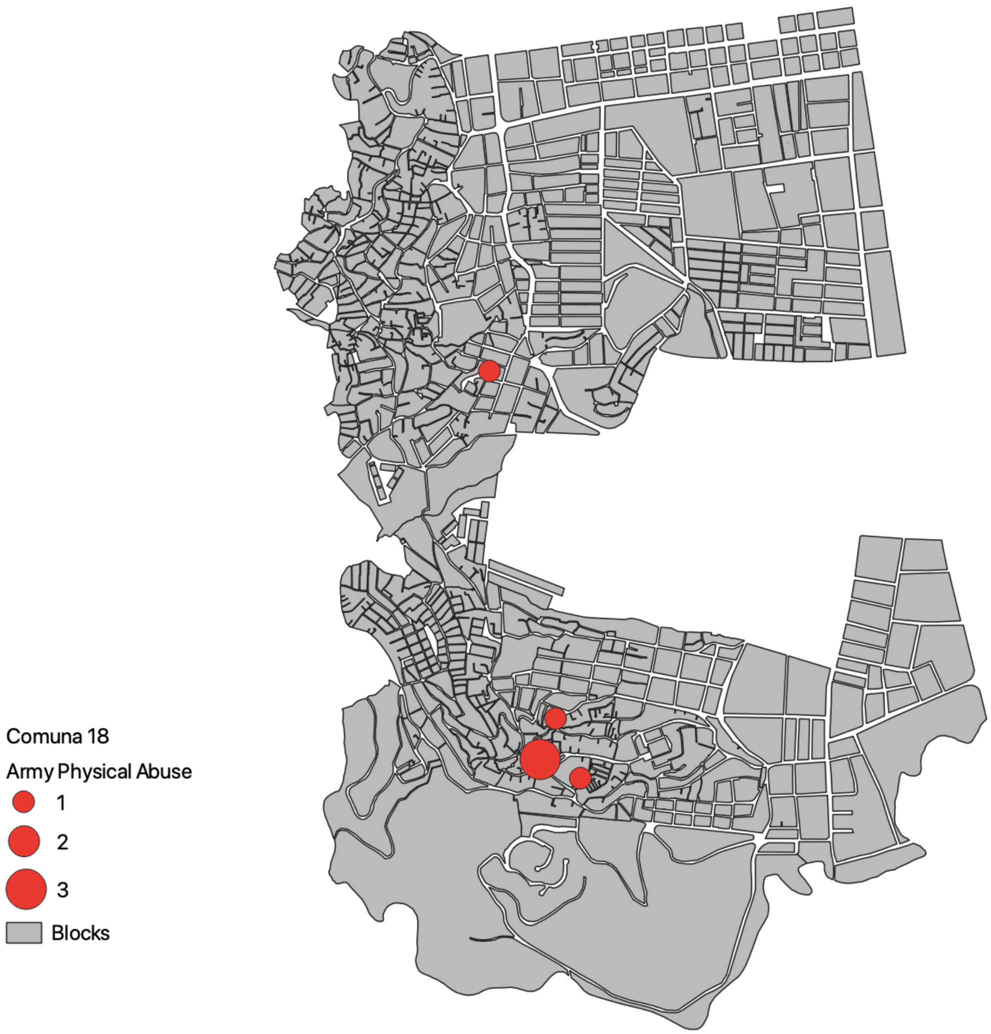

Notes: Red stars denote alleged abuses that occurred during the period of the intervention. Source: Instituto Geográfico Agustín Codazzi - Igac (underlying map) and Attorney-General's Office (abuse data).

Supplementary Figure 14: Alleged abuses by state security forces in commune 20 as reported to Attorney General's Office, September 30–November 18, 2019)




Notes: Red stars denote alleged abuses that occurred during the period of the intervention. Source: Instituto Geográfico Agustín Codazzi - Igac (underlying map) and Attorney-General's Office (abuse data).

Supplementary Figure 15: Military physical abuses in commune 18 in monitoring survey

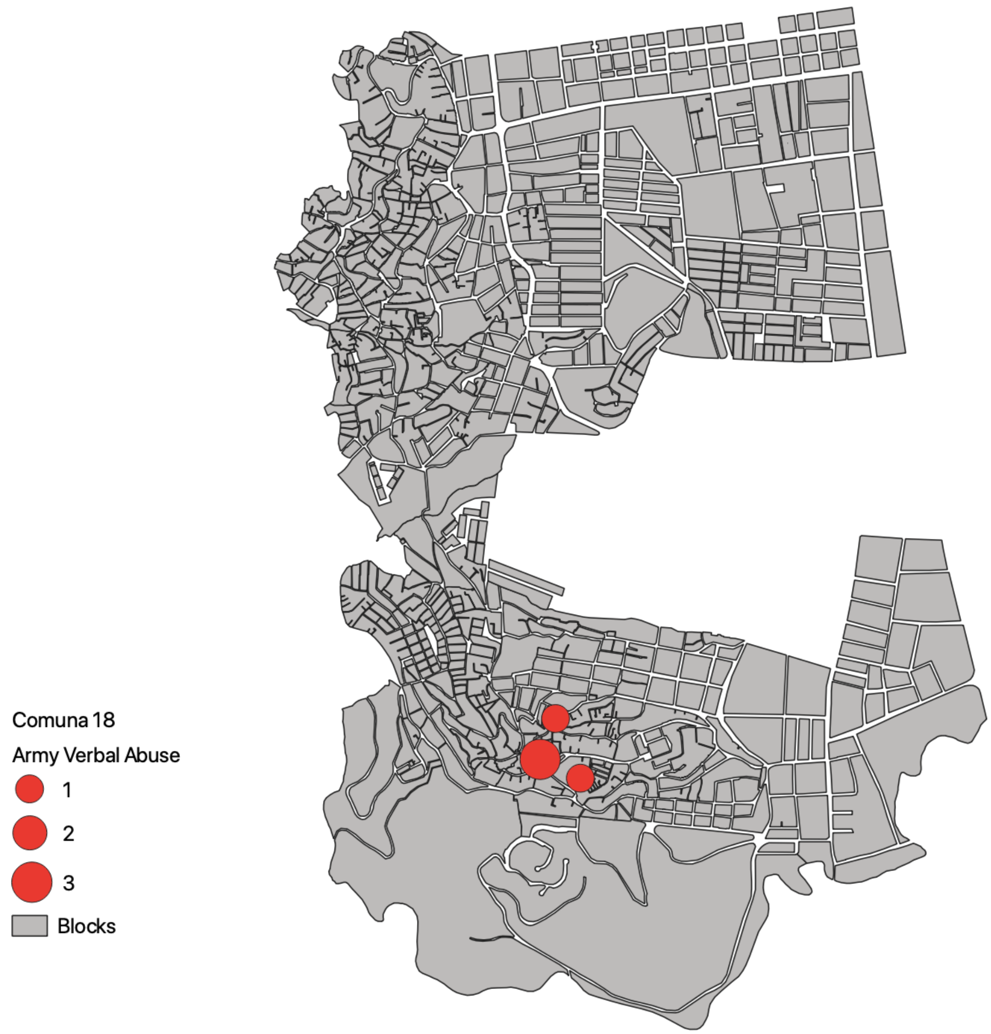



Notes: Red dots denote abuses. Source: Instituto Geográfico Agustín Codazzi - Igac (underlying map) and monitoring survey (abuse data).

Supplementary Figure 16: Military physical abuses in commune 20 in monitoring survey


Notes: Red dots denote abuses. Source: Instituto Geográfico Agustín Codazzi - Igac (underlying map) and monitoring survey (abuse data).

Supplementary Figure 17: Military verbal abuses in commune 18 in monitoring survey




Notes: Red dots denote abuses. Source: Instituto Geográfico Agustín Codazzi - Igac (underlying map) and monitoring survey (abuse data).

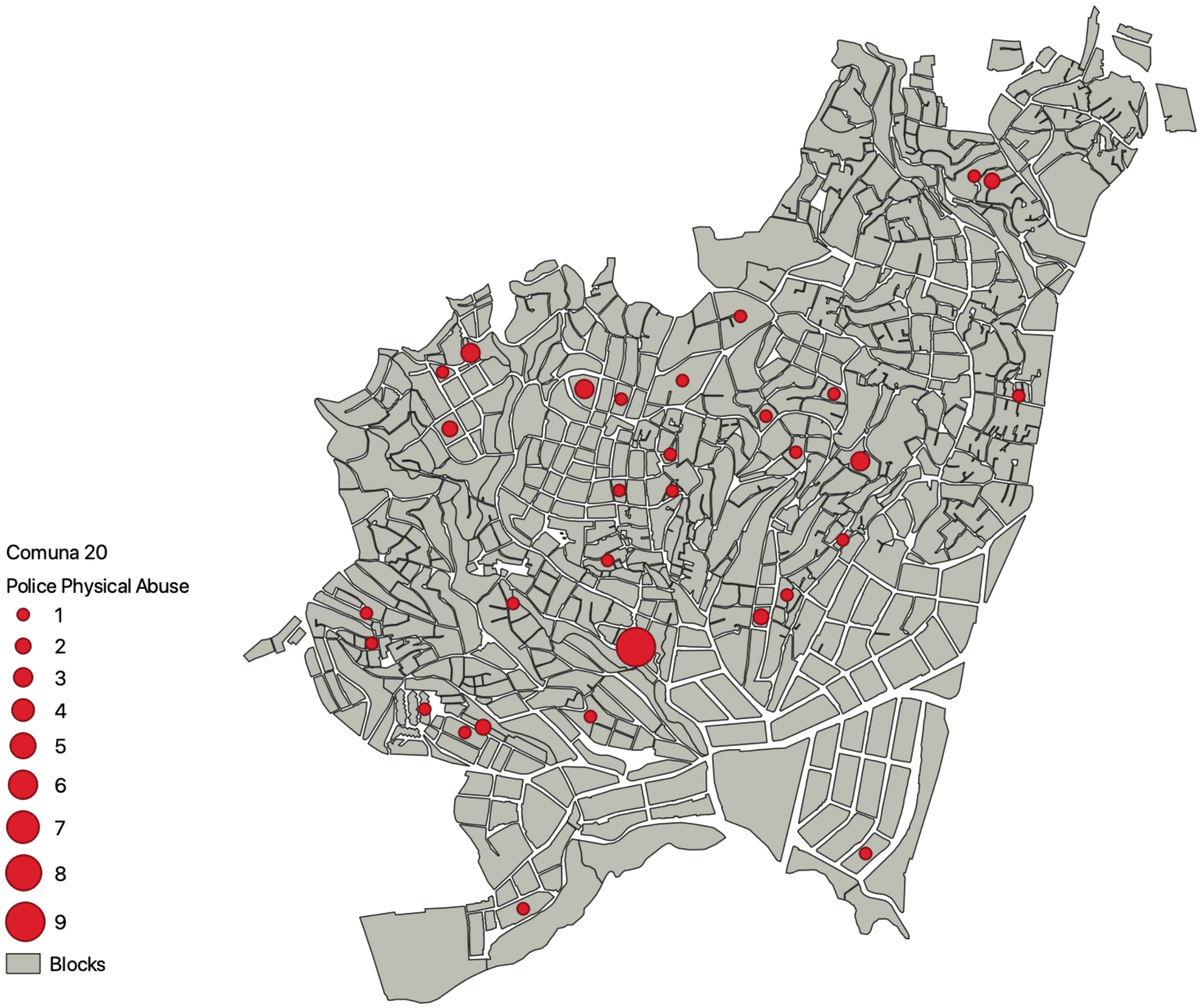
Supplementary Figure 18: Military physical abuses in commune 20 in monitoring survey


Notes: Red dots denote abuses. Source: Instituto Geográfico Agustín Codazzi - Igac (underlying map) and monitoring survey (abuse data).

Supplementary Figure 19: Police physical abuses in commune 18 in monitoring survey

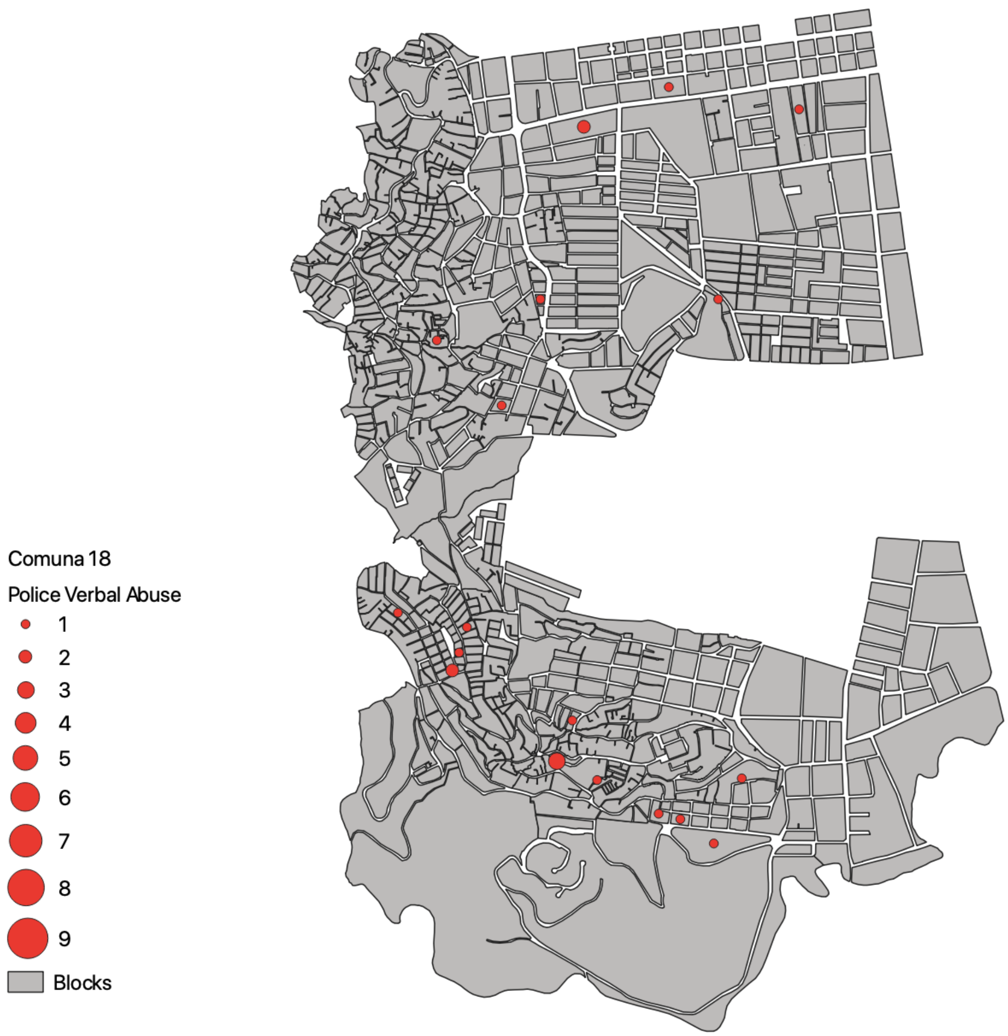

Notes: Red dots denote abuses. Source: Instituto Geográfico Agustín Codazzi - Igac (underlying map) and monitoring survey (abuse data).

Supplementary Figure 20: Police physical abuses in commune 20 in monitoring survey




Notes: Red dots denote abuses. Source: Instituto Geográfico Agustín Codazzi - Igac (underlying map) and monitoring survey (abuse data).

Supplementary Figure 21: Police verbal abuses in commune 18 in monitoring survey

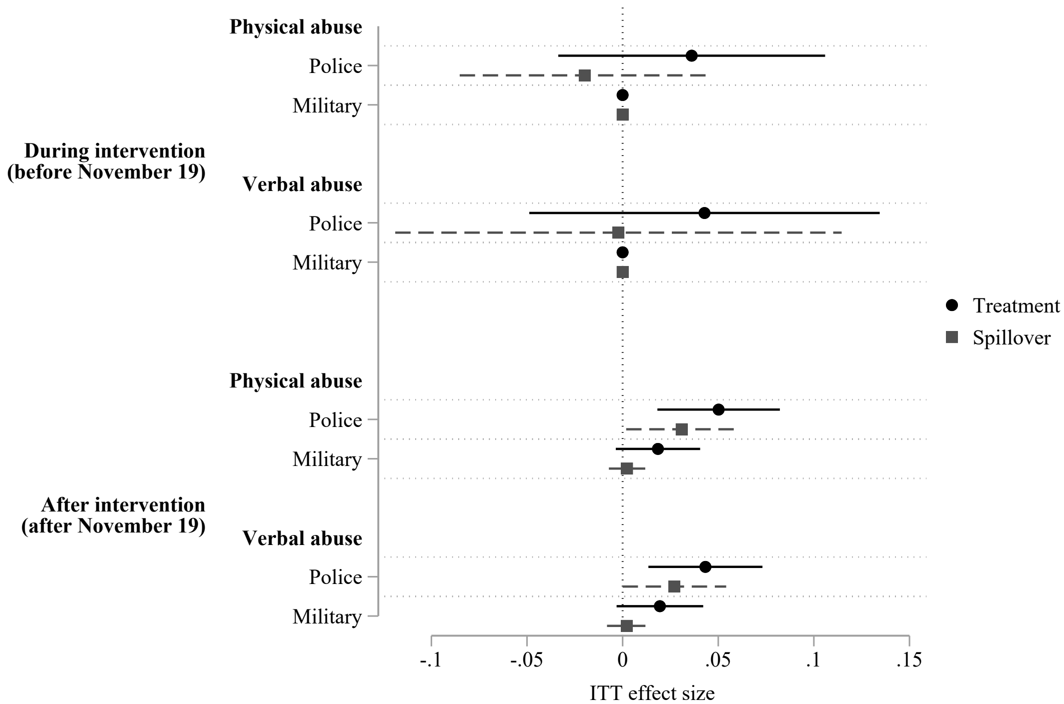



Notes: Red dots denote abuses. Source: Instituto Geográfico Agustín Codazzi - Igac (underlying map) and monitoring survey (abuse data).

Supplementary Figure 22: Police verbal abuses in commune 20 in monitoring survey

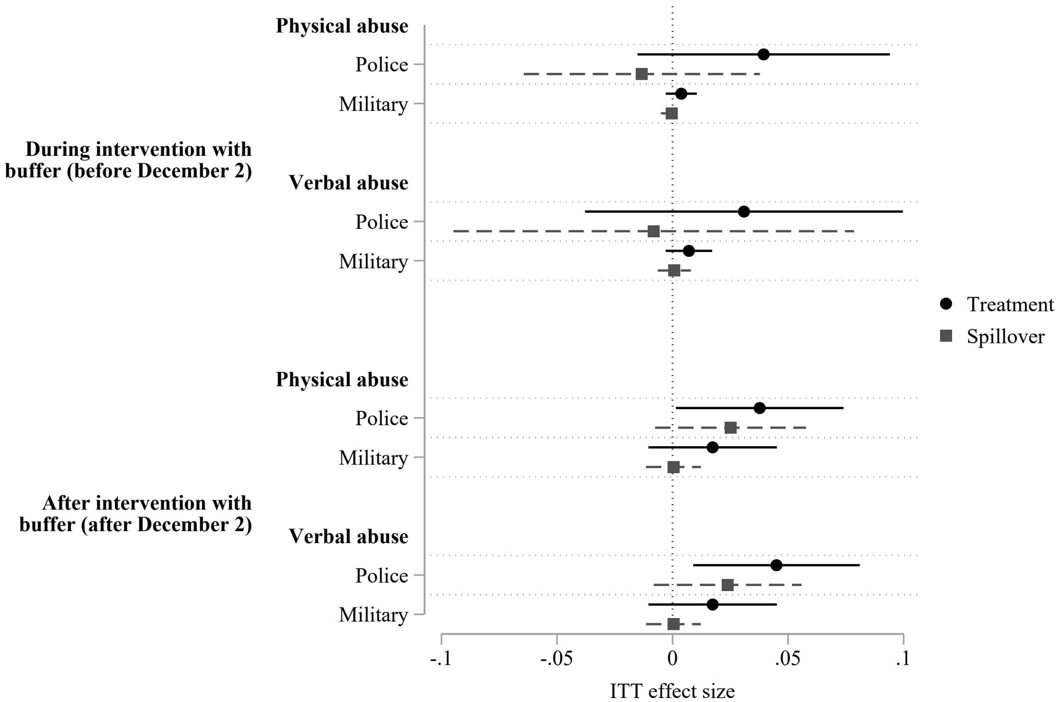

Notes: Red dots denote abuses. Source: Instituto Geográfico Agustín Codazzi - Igac (underlying map) and monitoring survey (abuse data).

Supplementary Figure 23: Treatment effects on abuses disaggregated by date of survey




Notes: ITT on abuses committed by police officers and soldiers based on monitoring data, distinguishing between abuses reported by survey respondents who were surveyed during the intervention (top panel) and abuses reported by survey respondents who were surveyed after the intervention (bottom panel). The dependent variable in all models is an indicator for any incidents of verbal or physical abuse on the block. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age and gender; models in the bottom panel include an additional individual-level control for years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "police physical abuse during intervention" ($N = 829$): $\beta = 0.036$, CI = [-0.034, 0.106], $p = 0.308$, RI $p = 0.653$ on treatment blocks; $\beta = -0.020$, CI = [-0.085, 0.046], $p = 0.552$, RI $p = 0.765$ on spillover blocks. For "police verbal abuse during intervention" ($N = 829$): $\beta = 0.043$, CI = [-0.049, 0.134], $p = 0.358$, RI $p = 0.637$ on treatment blocks; $\beta = -0.002$, CI = [-0.119, 0.115], $p = 0.970$, RI $p = 0.978$ on spillover blocks. For "police physical abuse after intervention" ($N = 1,141$): $\beta = 0.050$, CI = [0.018, 0.082], $p = 0.002$, RI $p = 0.198$ on treatment blocks; $\beta = 0.031$, CI = [0.002, 0.060], $p = 0.037$, RI $p = 0.198$ on spillover blocks. For "military physical abuse after intervention" ($N = 1,141$): $\beta = 0.018$, CI = [-0.004, 0.041], $p = 0.100$, RI $p = 0.465$ on treatment blocks; $\beta = 0.002$, CI = [-0.007, 0.012], $p = 0.631$, RI $p = 0.918$ on spillover blocks. For "police verbal abuse after intervention" ($N = 1,141$): $\beta = 0.043$, CI = [0.013, 0.073], $p = 0.005$, RI $p = 0.235$ on treatment blocks; $\beta = 0.027$, CI = [-0.000, 0.054], $p = 0.050$, RI $p = 0.364$ on spillover blocks. For "military verbal abuse after intervention" ($N = 1,141$): $\beta = 0.019$, CI = [-0.003, 0.042], $p = 0.091$, RI $p = 0.456$ on treatment blocks; $\beta = 0.002$, CI = [-0.008, 0.013], $p = 0.680$, RI $p = 0.933$ on spillover blocks. We do not report results for military abuse during the intervention because there were no incidents of physical or verbal abuse by soldiers reported by respondents surveyed before November 19.

Supplementary Figure 24: Treatment effects on abuses disaggregated by date of survey with buffer




Notes: ITT on abuses committed by police officers and soldiers based on monitoring data, distinguishing between abuses reported by survey respondents who were surveyed during or up to two weeks after the intervention (top panel) and abuses reported by survey respondents who were surveyed more than two weeks after the intervention (bottom panel). The dependent variable in all models is an indicator for any incidents of verbal or physical abuse on the block. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age and gender; models in the bottom panel include an additional individual-level control for years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "police physical abuse during intervention with buffer" ($N = 1,105$): $\beta = 0.039$, CI = [-0.015, 0.094], $p = 0.156$, RI $p = 0.513$ on treatment blocks; $\beta = -0.013$, CI = [-0.064, 0.038], $p = 0.608$, RI $p = 0.796$ on spillover blocks. For "military physical abuse during intervention with buffer" ($N = 1,105$): $\beta = 0.004$, CI = [-0.003, 0.011], $p = 0.267$, RI $p = 0.592$ on treatment blocks; $\beta = -0.000$, CI = [-0.005, 0.004], $p = 0.907$, RI $p = 0.974$ on spillover blocks. For "police verbal abuse during intervention with buffer" ($N = 1,105$): $\beta = 0.031$, CI = [-0.038, 0.100], $p = 0.377$, RI $p = 0.631$ on treatment blocks; $\beta = -0.008$, CI = [-0.095, 0.079], $p = 0.853$, RI $p = 0.897$ on spillover blocks. For "military verbal abuse during intervention with buffer" ($N = 1,105$): $\beta = 0.007$, CI = [-0.003, 0.017], $p = 0.165$, RI $p = 0.395$ on treatment blocks; $\beta = 0.001$, CI = [-0.006, 0.008], $p = 0.830$, RI $p = 0.926$ on spillover blocks. For "police physical abuse after intervention with buffer" ($N = 865$): $\beta = 0.038$, CI = [0.001, 0.074], $p = 0.042$, RI $p = 0.443$ on treatment blocks; $\beta = 0.025$, CI = [-0.008, 0.058], $p = 0.131$, RI $p = 0.554$ on spillover blocks. For "military physical abuse after intervention with buffer" ($N = 865$): $\beta = 0.017$, CI = [-0.010, 0.045], $p = 0.219$, RI $p = 0.579$ on treatment blocks; $\beta = 0.000$, CI = [-0.011, 0.012], $p = 0.947$, RI $p = 0.991$ on spillover blocks. For "police verbal abuse after intervention with buffer" ($N = 865$): $\beta = 0.045$, CI = [0.009, 0.081], $p = 0.015$, RI $p = 0.373$ on treatment blocks; $\beta = 0.024$, CI = [-0.008, 0.056], $p = 0.142$, RI $p = 0.585$ on spillover blocks. For "military verbal abuse after intervention with buffer" ($N = 865$): $\beta = 0.017$, CI = [-0.010, 0.045], $p = 0.219$, RI $p = 0.579$ on treatment blocks; $\beta = 0.000$, CI = [-0.011, 0.012], $p = 0.947$, RI $p = 0.991$ on spillover blocks.

Supplementary Table 14: Heterogeneous treatment effects on crime, crime victimization, and crime witnessing by prior crime rate

| | Admin data | | Endline survey | | |
|---|---|---|---|---|---|
| | Crime incidence | | Crime victimization | | Crime witnessing |
| | During intervention | After intervention | During intervention | After intervention | After intervention |
| Treatment | -0.002 | -0.053 | 0.049 | 0.027 | 0.184 |
| | [-0.075, 0.071] | [-0.142, 0.035] | [-0.040, 0.137] | [-0.068, 0.123] | [0.073, 0.294] |
| | (0.953) | (0.237) | (0.280) | (0.573) | (0.001) |
| Spillover | -0.020 | 0.003 | 0.056 | 0.048 | 0.218 |
| | [-0.084, 0.044] | [-0.070, 0.077] | [-0.006, 0.119] | [-0.028, 0.124] | [0.123, 0.313] |
| | (0.536) | (0.925) | (0.077) | (0.216) | (0.000) |
| Prior crime | 0.089 | -0.086 | 0.056 | 0.038 | 0.054 |
| | [0.008, 0.170] | [-0.327, 0.154] | [0.014, 0.099] | [-0.031, 0.107] | [0.001, 0.107] |
| | (0.031) | (0.481) | (0.010) | (0.276) | (0.047) |
| Treatment × Prior crime | 0.007 | 0.228 | -0.058 | -0.047 | -0.042 |
| | [-0.075, 0.088] | [0.160, 0.297] | [-0.101, -0.015] | [-0.116, 0.022] | [-0.095, 0.010] |
| | (0.874) | (0.000) | (0.008) | (0.179) | (0.116) |
| Spillover × Prior crime | -0.030 | 0.099 | -0.040 | -0.050 | -0.045 |
| | [-0.123, 0.063] | [0.010, 0.187] | [-0.088, 0.007] | [-0.122, 0.022] | [-0.103, 0.013] |
| | (0.531) | (0.028) | (0.098) | (0.176) | (0.132) |
| Individual controls | No | No | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes | Yes | Yes |
| Observations | 1167 | 1167 | 7845 | 7845 | 7837 |
| $R^2$ | 0.34 | 0.52 | 0.03 | 0.03 | 0.12 |
| Control mean | 0.160 | 0.160 | -0.021 | -0.016 | -0.119 |

Notes: HTEs 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. Treatment and spillover indicators are interacted with an additive index of crimes committed before the intervention on each block based on administrative data. The dependent variable in columns 1 and 2 is an additive index of murders, armed robberies, thefts, illegal drug sales, and illegal possession of weapons. The dependent variable in columns 3 and 4 is a standardized additive index of victimization by vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, domestic violence, or extortion. The dependent variable in column 5 is a standardized additive index of witnessing vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, domestic violence, extortion, prostitution, illegal drug sales, illegal drug consumption, public alcohol consumption, or illegal possession of firearms in the past month. The frequency of crime witnessing was measured on a 1-4 Likert scale. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. Models 1 and 2 also include a lagged dependent variable and block-level controls for the average age, average years of education, and percentage of men on each block. Models 3-5 include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors in models 3-5 are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Table 15: Heterogeneous treatment effects on crime victimization and crime witnessing by gender

| | Endline survey | | |
|---|---|---|---|
| | Crime victimization | | Crime witnessing |
| | During intervention | After intervention | During intervention |
| Treatment | -0.045 | -0.068 | 0.144 |
| | [-0.190, 0.100] | [-0.218, 0.082] | [-0.007, 0.296] |
| | (0.546) | (0.372) | (0.062) |
| Spillover | 0.029 | -0.042 | 0.134 |
| | [-0.092, 0.150] | [-0.160, 0.076] | [0.019, 0.250] |
| | (0.639) | (0.483) | (0.023) |
| Female | -0.042 | -0.066 | -0.041 |
| | [-0.167, 0.083] | [-0.176, 0.045] | [-0.145, 0.063] |
| | (0.513) | (0.242) | (0.442) |
| Treatment × Female | 0.076 | 0.092 | 0.014 |
| | [-0.089, 0.240] | [-0.067, 0.250] | [-0.144, 0.171] |
| | (0.366) | (0.258) | (0.865) |
| Spillover × Female | -0.005 | 0.082 | 0.077 |
| | [-0.147, 0.138] | [-0.044, 0.208] | [-0.042, 0.195] |
| | (0.950) | (0.201) | (0.204) |
| Individual controls | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes |
| Observations | 7845 | 7845 | 7837 |
| $R^2$ | 0.03 | 0.03 | 0.12 |
| Control mean | -0.021 | -0.016 | -0.119 |

Notes: HTEs 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. Treatment and spillover indicators are interacted with an indicator for gender. The dependent variable in columns 1 and 2 is a standardized additive index of victimization by vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, or extortion. The dependent variable in column 3 is a standardized additive index of witnessing vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, domestic violence, extortion, prostitution, illegal drug sales, illegal drug consumption, public alcohol consumption, or illegal possession of a firearm in the past month. The frequency of crime witnessing was measured on a 1-4 Likert scale. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Table 16: Heterogeneous treatment effects on perceptions of safety by prior crime rate

| | Endline survey | | |
|---|---|---|---|
| | **All safety** | **Personal safety** | **Business safety** |
| Treatment | -0.112 | -0.114 | 0.217 |
| | [-0.212, -0.013] | [-0.213, -0.015] | [-0.049, 0.483] |
| | (0.026) | (0.024) | (0.110) |
| Spillover | -0.130 | -0.133 | 0.017 |
| | [-0.216, -0.045] | [-0.219, -0.048] | [-0.230, 0.264] |
| | (0.003) | (0.002) | (0.893) |
| Prior crime | -0.083 | -0.083 | -0.065 |
| | [-0.129, -0.037] | [-0.129, -0.037] | [-0.170, 0.040] |
| | (<0.001) | (<0.001) | (0.224) |
| Treatment × Prior crime | 0.088 | 0.087 | 0.061 |
| | [0.041, 0.134] | [0.040, 0.133] | [-0.043, 0.166] |
| | (<0.001) | (<0.001) | (0.251) |
| Spillover × Prior crime | 0.092 | 0.093 | 0.068 |
| | [0.039, 0.145] | [0.039, 0.146] | [-0.045, 0.180] |
| | (0.001) | (0.001) | (0.237) |
| Individual controls | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes |
| Observations | 7707 | 7708 | 1041 |
| $R^2$ | 0.09 | 0.09 | 0.13 |
| Control mean | 0.077 | 0.078 | -0.065 |

Notes: HTEs on perceptions of safety for respondents and business owners together (column 1), residents only (column 2), and business owners only (column 3) based on survey data. Treatment and spillover indicators are interacted with an additive index of crimes committed before the intervention on each block based on administrative data. Residents and business owners were asked how safe they feel walking their blocks during the day and at night; how safe they feel talking on a smartphone on their block; and how worried they are about becoming victims of a violent or non-violent crime in the next two weeks. Perceptions of safety were measured on a 1-5 Likert scale. Residents and business owners were also asked whether they had taken precautions for fear of crime in the past month, including avoiding leaving their home or business at night; avoiding public transportation or recreation areas; prohibiting children from playing in the streets or attending school; considering moving to a different neighborhood; or changing jobs or schools. The dependent variable in models 1 and 2 is a standardized additive index based on responses to these questions. Business owners were also asked if they had changed their hours, hired private security guards, or closed their businesses for fear of crime. The dependent variable in model 3 is a standardized additive index based on responses to these questions. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Table 17: Heterogeneous treatment effects on perceptions of safety by gender

| | Endline survey | | |
|---|---|---|---|
| | **All safety** | **Personal safety** | **Business safety** |
| Treatment | -0.023 | -0.027 | 0.340 |
| | [-0.155, 0.110] | [-0.158, 0.105] | [0.002, 0.678] |
| | (0.739) | (0.691) | (0.049) |
| Spillover | -0.004 | -0.006 | 0.066 |
| | [-0.114, 0.106] | [-0.115, 0.104] | [-0.235, 0.367] |
| | (0.941) | (0.917) | (0.667) |
| Female | -0.166 | -0.166 | -0.078 |
| | [-0.275, -0.058] | [-0.274, -0.057] | [-0.425, 0.269] |
| | (0.003) | (0.003) | (0.659) |
| Treatment × Female | -0.041 | -0.038 | -0.096 |
| | [-0.192, 0.111] | [-0.189, 0.114] | [-0.516, 0.324] |
| | (0.600) | (0.625) | (0.654) |
| Spillover × Female | -0.092 | -0.093 | 0.043 |
| | [-0.215, 0.031] | [-0.216, 0.030] | [-0.341, 0.427] |
| | (0.143) | (0.139) | (0.826) |
| Individual controls | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes |
| Observations | 7707 | 7708 | 1041 |
| $R^2$ | 0.09 | 0.09 | 0.13 |
| Control mean | 0.077 | 0.078 | -0.065 |

Notes: HTEs on perceptions of safety for respondents and business owners together (column 1), residents only (column 2), and business owners only (column 3) based on survey data. Treatment and spillover indicators are interacted with an indicator for gender. Residents and business owners were asked how safe they feel walking their blocks during the day and at night; how safe they feel talking on a smartphone on their block; and how worried they are about becoming victims of a violent or non-violent crime in the next two weeks. Perceptions of safety were measured on a 1-5 Likert scale. Residents and business owners were also asked whether they had taken precautions for fear of crime in the past month, including avoiding leaving their home or business at night; avoiding public transportation or recreation areas; prohibiting children from playing in the streets or attending school; considering moving to a different neighborhood; or changing jobs or schools. The dependent variable in models 1 and 2 is a standardized additive index based on responses to these questions. Business owners were also asked if they had changed their hours, hired private security guards, or closed their businesses for fear of crime. The dependent variable in model 3 is a standardized additive index based on responses to these questions. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Table 18: Heterogeneous treatment effects on abuses by prior crime rate

| | Monitoring survey | | Endline survey | |
|---|---|---|---|---|
| | **Police abuse** | **Military abuse** | **Police abuse** | **Military abuse** |
| Treatment | 0.037 | 0.006 | 0.013 | -0.001 |
| | [0.015, 0.059] | [-0.003, 0.016] | [-0.016, 0.041] | [-0.010, 0.007] |
| | (0.001) | (0.195) | (0.375) | (0.746) |
| Spillover | 0.007 | 0.001 | 0.033 | 0.002 |
| | [-0.015, 0.029] | [-0.007, 0.009] | [0.007, 0.059] | [-0.005, 0.010] |
| | (0.522) | (0.832) | (0.013) | (0.548) |
| Prior crime | 0.003 | 0.000 | -0.000 | -0.000 |
| | [-0.006, 0.011] | [-0.002, 0.003] | [-0.014, 0.013] | [-0.005, 0.004] |
| | (0.519) | (0.803) | (0.971) | (0.831) |
| Treatment $\times$ Prior crime | 0.002 | 0.009 | -0.003 | -0.000 |
| | [-0.022, 0.025] | [-0.013, 0.030] | [-0.016, 0.011] | [-0.004, 0.004] |
| | (0.893) | (0.415) | (0.670) | (0.993) |
| Spillover $\times$ Prior crime | 0.022 | 0.001 | -0.005 | -0.001 |
| | [-0.004, 0.048] | [-0.005, 0.007] | [-0.019, 0.009] | [-0.005, 0.004] |
| | (0.092) | (0.709) | (0.512) | (0.816) |
| Individual controls | Yes | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes | Yes |
| Block-level controls | No | No | No | No |
| Observations | 1970 | 1970 | 7908 | 7908 |
| $R^2$ | 0.05 | 0.05 | 0.05 | 0.01 |
| Control mean | 0.015 | 0.000 | 0.114 | 0.012 |

Notes: HTEs on abuses committed by police officers and soldiers based on monitoring (columns 1 and 2) and survey data (columns 3 and 4). Treatment and spillover indicators are interacted with an additive index of crimes committed before the intervention on each block based on administrative data. The dependent variable in all models is an indicator for any incidents of verbal or physical abuse on the block. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age and gender; models in columns 3 and 4 include an additional individual-level control for years of education. ITT effect sizes are derived from WLS regressions. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Table 19: Heterogeneous treatment effects on abuses by gender

| | **Monitoring survey** | | **Endline survey** | |
|---|---|---|---|---|
| | **Police abuse** | **Military abuse** | **Police abuse** | **Military abuse** |
| Treatment | 0.027 | 0.009 | 0.009 | 0.002 |
| | [-0.002, 0.056] | [-0.007, 0.024] | [-0.036, 0.053] | [-0.009, 0.014] |
| | (0.066) | (0.279) | (0.696) | (0.688) |
| Spillover | 0.010 | 0.007 | 0.035 | 0.008 |
| | [-0.023, 0.043] | [-0.006, 0.019] | [-0.004, 0.074] | [-0.001, 0.018] |
| | (0.554) | (0.298) | (0.075) | (0.092) |
| Female | -0.004 | -0.000 | -0.012 | 0.007 |
| | [-0.021, 0.014] | [-0.005, 0.004] | [-0.050, 0.027] | [-0.004, 0.017] |
| | (0.672) | (0.926) | (0.545) | (0.211) |
| Treatment $\times$ Female | 0.015 | 0.003 | 0.003 | -0.006 |
| | [-0.016, 0.045] | [-0.010, 0.016] | [-0.048, 0.054] | [-0.021, 0.009] |
| | (0.345) | (0.684) | (0.911) | (0.457) |
| Spillover $\times$ Female | 0.009 | -0.008 | -0.008 | -0.009 |
| | [-0.029, 0.047] | [-0.022, 0.005] | [-0.051, 0.036] | [-0.022, 0.003] |
| | (0.647) | (0.209) | (0.733) | (0.148) |
| Individual controls | Yes | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes | Yes |
| Block-level controls | No | No | No | No |
| Observations | 1970 | 1970 | 7908 | 7908 |
| $R^2$ | 0.05 | 0.05 | 0.05 | 0.01 |
| Control mean | 0.015 | 0.000 | 0.114 | 0.012 |

Notes: HTEs on abuses committed by police officers and soldiers based on monitoring (columns 1 and 2) and survey data (columns 3 and 4). Treatment and spillover indicators are interacted with an indicator for gender. The dependent variable in all models is an indicator for any incidents of verbal or physical abuse on the block. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age and gender; models in columns 3 and 4 include an additional individual-level control for years of education. ITT effect sizes are derived from WLS regressions. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Table 20: Heterogeneous treatment effects on crime victimization and crime witnessing by prior crime victimization

| | **Endline survey** | | |
|---|---|---|---|
| | **Crime victimization** | | **Crime witnessing** |
| | During intervention | After intervention | During intervention |
| Treatment | 0.004 | -0.008 | 0.146 |
| | [-0.074, 0.082] | [-0.098, 0.083] | [0.048, 0.244] |
| | (0.924) | (0.865) | (0.004) |
| Spillover | 0.025 | 0.013 | 0.182 |
| | [-0.032, 0.083] | [-0.060, 0.086] | [0.100, 0.264] |
| | (0.390) | (0.729) | (<0.001) |
| Prior crime victimization | 0.145 | 0.057 | 0.177 |
| | [0.057, 0.233] | [-0.028, 0.142] | [0.129, 0.225] |
| | (0.001) | (0.191) | (<0.001) |
| Treatment × Prior crime victimization | 0.042 | 0.078 | 0.046 |
| | [-0.117, 0.200] | [-0.062, 0.218] | [-0.031, 0.123] |
| | (0.608) | (0.276) | (0.243) |
| Spillover × Prior crime victimization | -0.005 | -0.004 | -0.052 |
| | [-0.109, 0.099] | [-0.098, 0.090] | [-0.108, 0.005] |
| | (0.922) | (0.938) | (0.073) |
| Individual controls | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes |
| Observations | 7845 | 7845 | 7800 |
| $R^2$ | 0.05 | 0.04 | 0.15 |
| Control mean | -0.021 | -0.016 | -0.119 |

Notes: HTEs 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. Treatment and spillover indicators are interacted with a standardized additive index of crime victimization before the intervention. The dependent variable in columns 1 and 2 is a standardized additive index of victimization by vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, or extortion. The dependent variable in column 3 is a standardized additive index of witnessing vandalism, armed robbery, burglary, theft, motor vehicle theft, homicide, attempted homicide, gang activity, domestic violence, extortion, prostitution, illegal drug sales, illegal drug consumption, public alcohol consumption, or illegal possession of a firearm in the past month. The frequency of crime witnessing was measured on a 1-4 Likert scale. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Table 21: Heterogeneous treatment effects on perceptions of safety by prior crime victimization

| | Endline survey | | |
|---|---|---|---|
| | **All safety** | **Personal safety** | **Business safety** |
| Treatment | -0.040 | -0.042 | 0.308 |
| | [-0.130, 0.050] | [-0.132, 0.048] | [0.091, 0.525] |
| | (0.382) | (0.357) | (0.006) |
| Spillover | -0.063 | -0.065 | 0.083 |
| | [-0.140, 0.013] | [-0.142, 0.011] | [-0.114, 0.280] |
| | (0.106) | (0.095) | (0.408) |
| Prior crime victimization | -0.144 | -0.142 | -0.177 |
| | [-0.206, -0.082] | [-0.204, -0.080] | [-0.311, -0.043] |
| | (<0.001) | (<0.001) | (0.010) |
| Treatment × Prior crime victimization | 0.008 | 0.004 | 0.060 |
| | [-0.070, 0.085] | [-0.074, 0.081] | [-0.156, 0.276] |
| | (0.848) | (0.921) | (0.585) |
| Spillover × Prior crime victimization | 0.027 | 0.025 | 0.096 |
| | [-0.042, 0.096] | [-0.044, 0.094] | [-0.058, 0.250] |
| | (0.444) | (0.474) | (0.223) |
| Individual controls | Yes | Yes | Yes |
| Neighborhood FE | Yes | Yes | Yes |
| Block-level controls | Yes | Yes | Yes |
| Observations | 7671 | 7672 | 1031 |
| $R^2$ | 0.11 | 0.11 | 0.15 |
| Control mean | 0.077 | 0.078 | -0.065 |

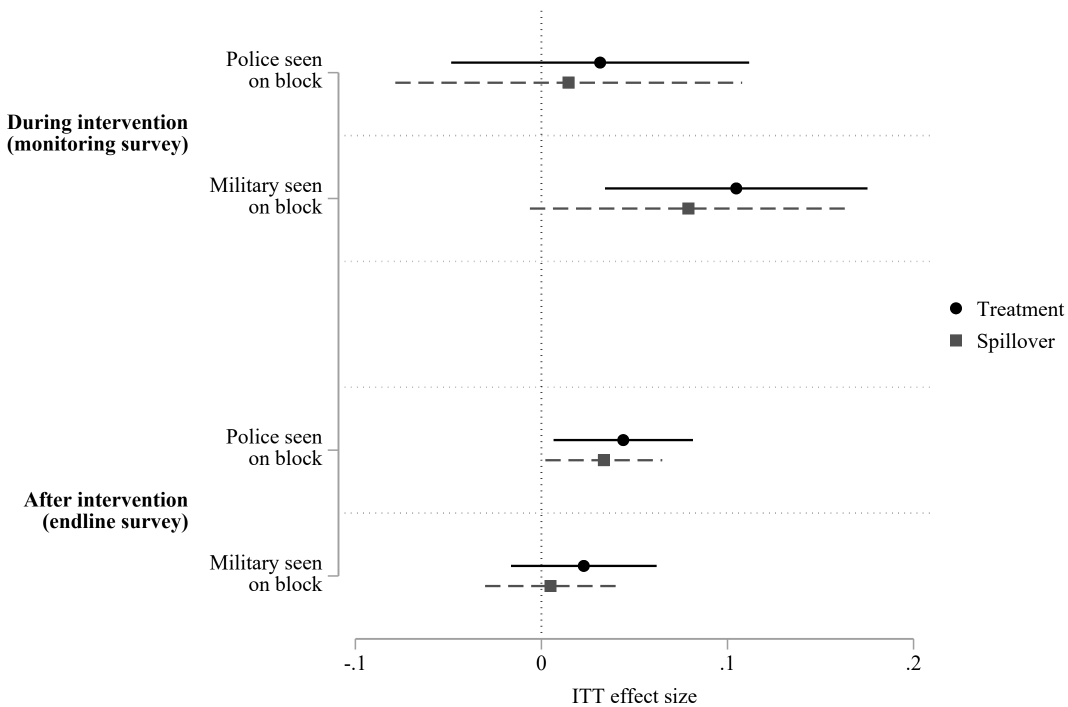
Notes: HTEs on perceptions of safety for respondents and business owners together (column 1), residents only (column 2), and business owners only (column 3) based on survey data. Treatment and spillover indicators are interacted with a standardized additive index of crime victimization before the intervention. Residents and business owners were asked how safe they feel walking their blocks during the day and at night; how safe they feel talking on a smartphone on their block; and how worried they are about becoming victims of a violent or non-violent crime in the next two weeks. Perceptions of safety were measured on a 1-5 Likert scale. Residents and business owners were also asked whether they had taken precautions for fear of crime in the past month, including avoiding leaving their home or business at night; avoiding public transportation or recreation areas; prohibiting children from playing in the streets or attending school; considering moving to a different neighborhood; or changing jobs or schools. The dependent variable in models 1 and 2 is a standardized additive index based on responses to these questions. Business owners were also asked if they had changed their hours, hired private security guards, or closed their businesses for fear of crime. The dependent variable in model 3 is a standardized additive index based on responses to these questions. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regression. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Table 22: Heterogeneous treatment effects on abuses by prior crime victimization

| | Endline survey | |
|---|---|---|
| | **Police abuse** | **Military abuse** |
| Treatment | 0.009 | -0.003 |
| | [-0.017, 0.035] | [-0.010, 0.005] |
| | (0.505) | (0.458) |
| Spillover | 0.030 | 0.002 |
| | [0.007, 0.053] | [-0.005, 0.008] |
| | (0.011) | (0.626) |
| Prior crime victimization | 0.025 | 0.006 |
| | [0.002, 0.047] | [-0.004, 0.016] |
| | (0.030) | (0.259) |
| Treatment × Prior crime victimization | 0.005 | 0.004 |
| | [-0.026, 0.037] | [-0.012, 0.020] |
| | (0.741) | (0.660) |
| Spillover × Prior crime victimization | 0.004 | -0.002 |
| | [-0.022, 0.030] | [-0.013, 0.010] |
| | (0.748) | (0.751) |
| Individual controls | Yes | Yes |
| Neighborhood FE | Yes | Yes |
| Block-level controls | Yes | Yes |
| Observations | 7870 | 7870 |
| $R^2$ | 0.05 | 0.02 |
| Control mean | 0.114 | 0.012 |

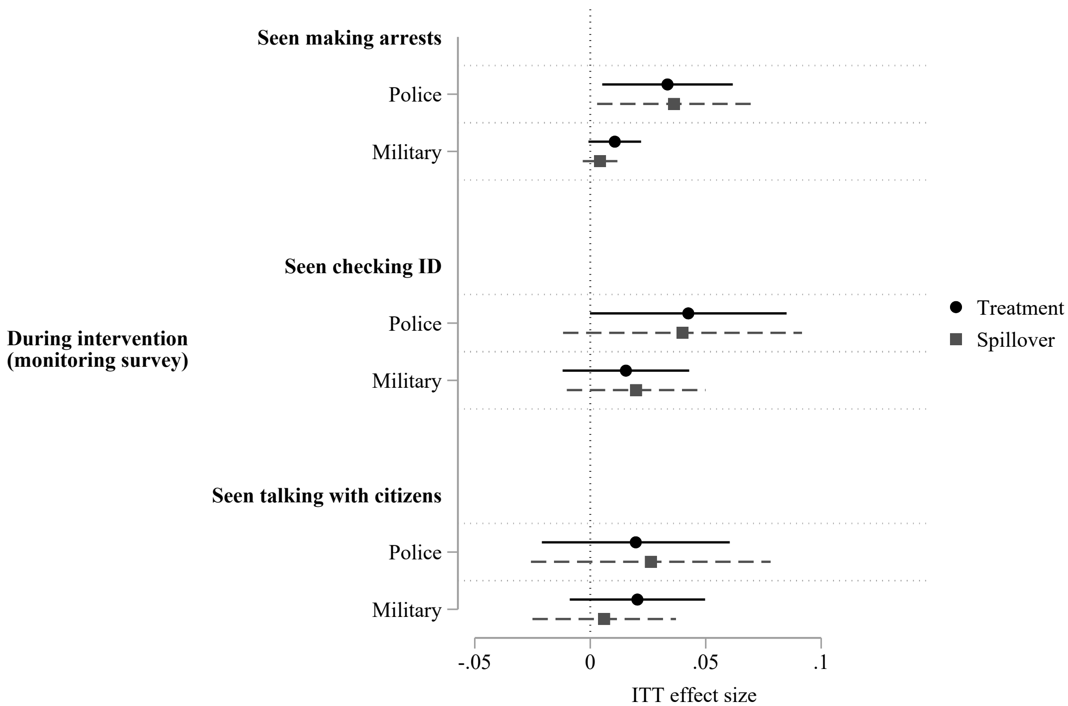
Notes: HTEs on abuses committed by police officers and soldiers based on monitoring (columns 1 and 2) and survey data (columns 3 and 4). Treatment and spillover indicators are interacted with a standardized additive index of crime victimization before the intervention. The dependent variable in all models is an indicator for any incidents of verbal or physical abuse on the block. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age and gender; models in columns 3 and 4 include an additional individual-level control for years of education. ITT effect sizes are derived from WLS regressions. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. 95% confidence intervals from two-tailed tests are in brackets. $p$-values are in parentheses. No adjustments were made for multiple comparisons.

Supplementary Figure 25: Treatment effects on exposure to police and military



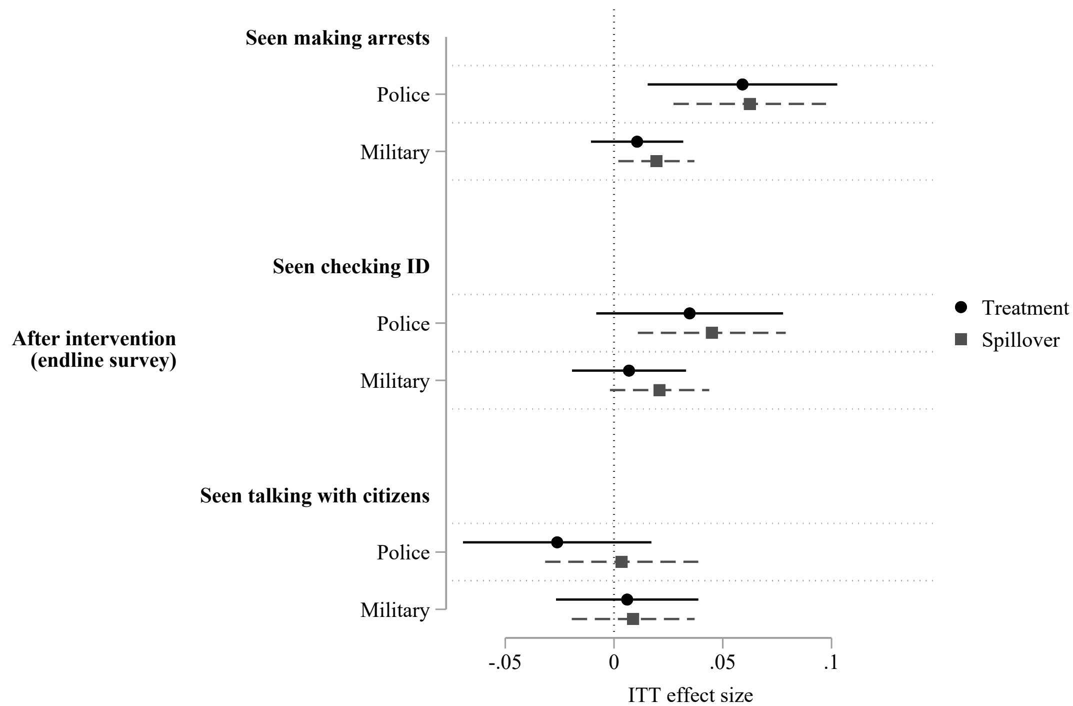

Notes: ITT on police and military presence during (top panel) and after (bottom panel) the intervention based on survey data. The dependent variables in the top panel are indicators for respondents' reports of any police or military presence on the block in the past two weeks. The dependent variables in the bottom panel are indicators for respondents' reports of somewhat or very frequent police or military presence on the block in the past month. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age and gender; models in the bottom panel include an additional individual-level control for years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "police seen on block during intervention" ($N = 1,970$): $\beta = 0.032$, CI = [-0.049, 0.112], $p = 0.439$, RI $p = 0.611$ on treatment blocks; $\beta = 0.015$, CI = [-0.079, 0.108], $p = 0.757$, RI $p = 0.771$ on spillover blocks. For "military seen on block during intervention" ($N = 1,970$): $\beta = 0.105$, CI = [0.034, 0.175], $p = 0.004$, RI $p = 0.064$ on treatment blocks; $\beta = 0.079$, CI = [-0.006, 0.164], $p = 0.069$, RI $p = 0.063$ on spillover blocks. For "police seen on block after intervention" ($N = 7,908$): $\beta = 0.044$, CI = [0.007, 0.081], $p = 0.021$, RI $p = 0.099$ on treatment blocks; $\beta = 0.034$, CI = [0.002, 0.065], $p = 0.037$, RI $p = 0.109$ on spillover blocks. For "military seen on block after intervention " ($N = 7,908$): $\beta = 0.023$, CI = [-0.016, 0.062], $p = 0.254$, RI $p = 0.425$ on treatment blocks; $\beta = 0.005$, CI = [-0.030, 0.040], $p = 0.786$, RI $p = 0.829$ on spillover blocks.

Supplementary Figure 26: Treatment effects on police and military activities during intervention



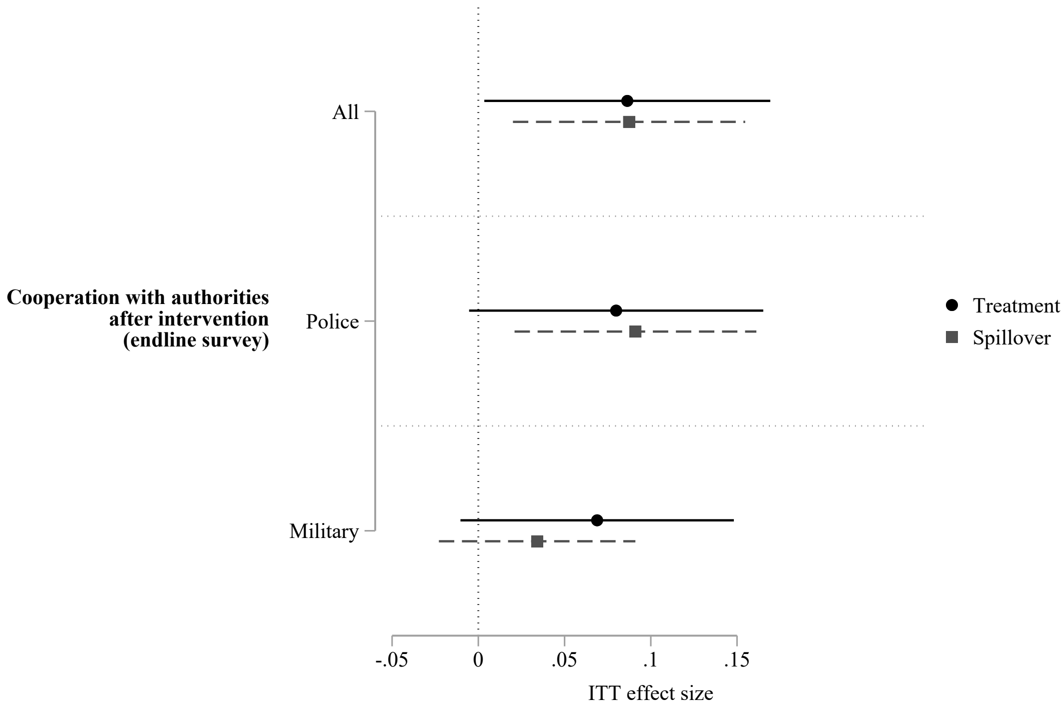

Notes: ITT on police and military presence during the intervention based on survey data. The dependent variables are indicators for respondents' reports of any police officers or soldiers making arrests, checking IDs, or talking with citizens on the block in the past two weeks. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age and gender. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "police seen making arrests" ($N = 1,970$): $\beta = 0.033$, CI = [0.005, 0.062], $p = 0.021$, RI $p = 0.210$ on treatment blocks; $\beta = 0.036$, CI = [0.003, 0.070], $p = 0.033$, RI $p = 0.082$ on spillover blocks. For "military seen making arrests" ($N = 1,970$): $\beta = 0.011$, CI = [-0.001, 0.022], $p = 0.068$, RI $p = 0.420$ on treatment blocks; $\beta = 0.004$, CI = [-0.003, 0.012], $p = 0.265$, RI $p = 0.721$ on spillover blocks. For "police seen checking ID" ($N = 1,970$): $\beta = 0.042$, CI = [-0.000, 0.085], $p = 0.051$, RI $p = 0.272$ on treatment blocks; $\beta = 0.040$, CI = [-0.012, 0.092], $p = 0.130$, RI $p = 0.174$ on spillover blocks. For "military seen checking ID" ($N = 1,970$): $\beta = 0.015$, CI = [-0.012, 0.043], $p = 0.269$, RI $p = 0.487$ on treatment blocks; $\beta = 0.020$, CI = [-0.010, 0.050], $p = 0.194$, RI $p = 0.244$ on spillover blocks. For "police seen talking with citizens" ($N = 1,970$): $\beta = 0.020$, CI = [-0.021, 0.060], $p = 0.341$, RI $p = 0.598$ on treatment blocks; $\beta = 0.026$, CI = [-0.026, 0.078], $p = 0.321$, RI $p = 0.367$ on spillover blocks. For "military seen talking with citizens" ($N = 1,970$): $\beta = 0.020$, CI = [-0.009, 0.050], $p = 0.171$, RI $p = 0.385$ on treatment blocks; $\beta = 0.006$, CI = [-0.025, 0.037], $p = 0.702$, RI $p = 0.735$ on spillover blocks.

Supplementary Figure 27: Treatment effects on police and military activities after intervention




Notes: ITT on police and military presence after the intervention based on survey data. The dependent variables are indicators for respondents' reports of police officers or soldiers making arrests, checking IDs, or talking with citizens somewhat or very frequently on the block in the past month. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "police seen making arrests" ($N = 7,908$): $\beta = 0.059$, CI = [0.015, 0.103], $p = 0.008$, RI $p = 0.068$ on treatment blocks; $\beta = 0.062$, CI = [0.027, 0.097], $p = 0.000$, RI $p = 0.013$ on spillover blocks. For "military seen making arrests" ($N = 7,908$): $\beta = 0.011$, CI = [-0.011, 0.032], $p = 0.327$, RI $p = 0.497$ on treatment blocks; $\beta = 0.019$, CI = [0.002, 0.037], $p = 0.029$, RI $p = 0.105$ on spillover blocks. For "police seen checking ID" ($N = 7,908$): $\beta = 0.035$, CI = [-0.008, 0.078], $p = 0.112$, RI $p = 0.306$ on treatment blocks; $\beta = 0.045$, CI = [0.011, 0.079], $p = 0.010$, RI $p = 0.090$ on spillover blocks. For "military seen checking ID" ($N = 7,908$): $\beta = 0.007$, CI = [-0.019, 0.033], $p = 0.606$, RI $p = 0.732$ on treatment blocks; $\beta = 0.021$, CI = [-0.002, 0.044], $p = 0.072$, RI $p = 0.174$ on spillover blocks. For "police seen talking with citizens" ($N = 7,908$): $\beta = -0.026$, CI = [-0.069, 0.017], $p = 0.237$, RI $p = 0.413$ on treatment blocks; $\beta = 0.004$, CI = [-0.032, 0.039], $p = 0.844$, RI $p = 0.885$ on spillover blocks. For "military seen talking with citizens" ($N = 7,908$): $\beta = 0.006$, CI = [-0.027, 0.039], $p = 0.717$, RI $p = 0.802$ on treatment blocks; $\beta = 0.009$, CI = [-0.019, 0.037], $p = 0.541$, RI $p = 0.645$ on spillover blocks.

Supplementary Figure 28: Treatment effects on cooperation with police and military




Notes: ITT on cooperation with the police and military after the intervention based on survey data. The dependent variables are standardized additive indices capturing respondents' reports that someone on their block had contacted the police or military to alert them to suspicious or criminal activity or to assist with a criminal investigation in the past month. All models include neighborhood fixed effects and block-level controls for the area of each block and the distance to the nearest police station, military battalion, and public transportation hub. All models also include individual-level controls for age, gender, and years of education. ITT effect sizes are derived from WLS regressions as in equation 2. Observations are weighted by the inverse probability of assignment to their realized treatment status. Standard errors are clustered by block. Circles and squares denote the treatment and spillover ITTs, respectively. Solid and dashed lines denote 95% confidence intervals from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For "all" ($N = 7,824$): $\beta = 0.086$, CI = [0.003, 0.169], $p = 0.041$, RI $p = 0.148$ on treatment blocks; $\beta = 0.087$, CI = [0.020, 0.155], $p = 0.011$, RI $p = 0.061$ on spillover blocks. For "police" ($N = 7,847$): $\beta = 0.080$, CI = [-0.005, 0.165], $p = 0.066$, RI $p = 0.196$ on treatment blocks; $\beta = 0.091$, CI = [0.021, 0.161], $p = 0.011$, RI $p = 0.059$ on spillover blocks. For "military" ($N = 7,866$): $\beta = 0.069$, CI = [-0.010, 0.148], $p = 0.088$, RI $p = 0.228$ on treatment blocks; $\beta = 0.034$, CI = [-0.023, 0.091], $p = 0.240$, RI $p = 0.446$ on spillover blocks.