nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01600-1

Little evidence that military policing reduces crime or improves human security

Received: 28 June 2021

Accepted: 12 April 2023

Published online: 11 May 2023

Check for updates

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

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. Here 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 probably 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 probably small and not worth the costs.

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 United States, commentators occasionally urge the deployment of troops to support police departments in 'high-crime, drug-infested urban areas' (ref. 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 programme in Cali, Colombia. The programme involved recurring, intensive military patrols targeting hot spots for crime. Military policing programmes 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 (for a summary of this debate, see ref. 24).

But these arguments are hotly contested. Opponents of military policing counter that deterrence depends on the ability to investigate

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

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].

So far, 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 Special Weapons And Tactics (SWAT) team deployments or transfers of military hardware to police departments in the United States, 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 (ref. 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.

In this article, 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 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 programme, 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 programme and our evaluation of it in the Methods section.

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 ongoing, and if anything our results suggest that it exacerbated crime after the intervention was complete. We observe an increase in post-intervention crime in the administrative data 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 artefacts of heightened vigilance on the part of civilians.

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 do not support the claims of military policing's proponents, and 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 programmes in other contexts. As in many Latin American cities, access to security and other services in Cali depends on where one lives and the socio-economic 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 programme we evaluate is place-based, analogous to hot spots policing interventions in the United States and elsewhere, including Colombia[1,40]. Almost all studies of hot spots policing focus on particular neighbourhoods 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 Fig. 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.

Results

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 programme 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 the Methods section.

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 ($\beta = 0.003$, 95% confidence interval (CI) −0.068 to 0.074, $P = 0.934$, RI $P = 0.959$, BF_{10} 0.107). This null effect is unlikely to be an artefact 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

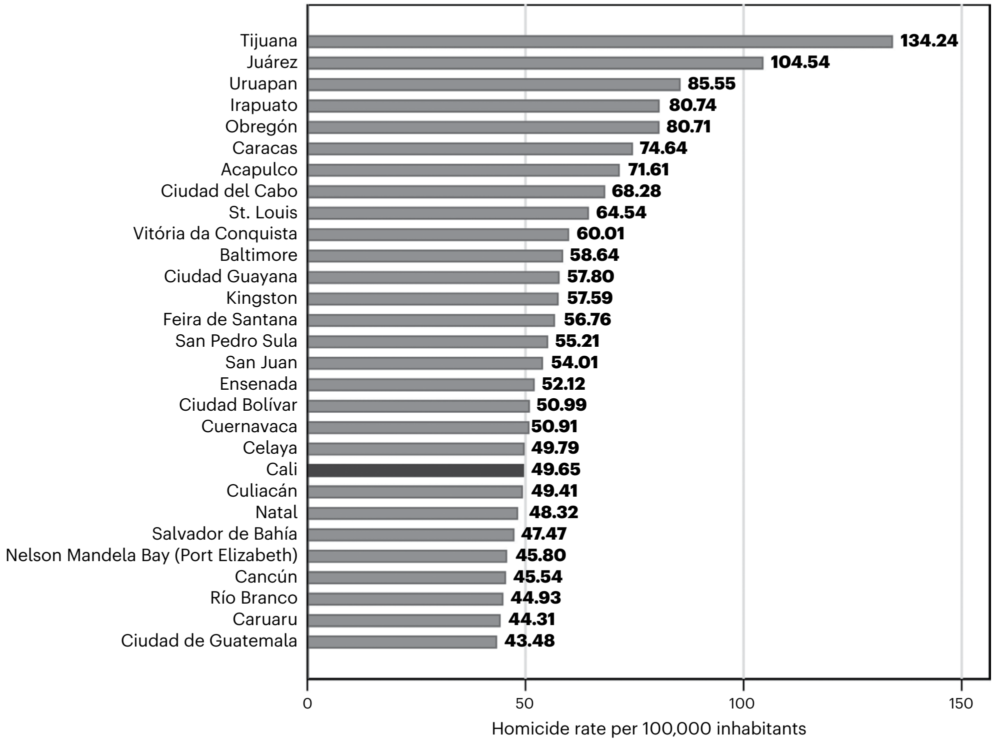

Fig. 1 | The world's most dangerous cities by homicide rate, 2019. Data on homicides is from Seguridad, Justicia y Paz, a non-governmental organization.

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 to 0.074	0.011 to 0.208	−0.077 to 0.089	−0.098 to 0.085	0.051 to 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 to 0.022	−0.003 to 0.169	−0.034 to 0.086	−0.061 to 0.087	0.101 to 0.270
	(0.212)	(0.059)	(0.389)	(0.729)	(<0.001)
Individual controls	No	No	Yes	Yes	Yes
Neighbourhood FE	Yes	Yes	Yes	Yes	Yes
Block-level controls	Yes	Yes	Yes	Yes	Yes
Observations	1,167	1,167	7,845	7,845	7,837
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 endline survey data; and crime witnessing after the intervention (column 5) based on endline 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 neighbourhood fixed effects (FE) 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. The 95% CIs are given as range, and *P* values are in parentheses. RI *P* values are in the last two rows. No adjustments were made for multiple comparisons; results with multiple comparisons corrections are reported in Supplementary Tables 6 and 7.

the intervention was ongoing (β = −0.038, 95% CI −0.097 to 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 19 November 2019 and the end of the year (β = 0.110, 95% CI 0.011 to 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 (β = 0.083, 95% CI −0.003 to 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 ten crimes in the past 6 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 the Methods section 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 (β = 0.006, 95% CI −0.077 to 0.089, P = 0.886, RI P = 0.927, BF_{10} 0.045) or spillover blocks (β = 0.026, 95% CI −0.034 to 0.086, P = 0.389, RI P = 0.610, BF_{10} 0.055) during the intervention, or on treatment (β = −0.007, 95% CI −0.098 to 0.085, P = 0.886, RI P = 0.914, BF_{10} 0.042) or spillover blocks (β = 0.013, 95% CI −0.061 to 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 favour 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 section.

Finally, Table 1 reports the ITT on residents' reports of witnessing crime in the endline survey (column 5). In addition to the ten 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 (β = 0.153, 95% CI 0.051 to 0.256, P = 0.003, RI P = 0.038) and by 0.186 standard deviations on spillover blocks (β = 0.186, 95% CI 0.101 to 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 police on spillover blocks (β = 0.091, 95% CI 0.021 to 0.161, P = 0.011, RI P = 0.059); respondents were also more likely to report crimes on treatment blocks, though this ITT is not quite statistically significant at conventional levels (β = 0.080, 95% CI −0.005 to 0.165, P = 0.066, RI P = 0.196). We also show that endline survey respondents were more likely to observe police officers making arrests on both treatment (β = 0.059, 95% CI 0.015 to 0.103, P = 0.008, RI P = 0.068) and spillover blocks (β = 0.062, 95% CI 0.027 to 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.

Crime disaggregated by day and time

All Plan Fortaleza patrols occurred on weekday nights. Following our PAP, in Fig. 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 (β = 0.006, 95% CI −0.053 to 0.066, P = 0.834, RI P = 0.901, BF_{10} 0.104) and at night (β = −0.008, 95% CI −0.054 to 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 (β = −0.038, 95% CI −0.088 to 0.011, P = 0.129, RI P = 0.365, BF_{10} 0.907) or at night (β = −0.017, 95% CI −0.060 to 0.025, P = 0.424, RI P = 0.514, BF_{10} 0.175).

Perceptions of safety

Sceptics of place-based policing worry that citizens will interpret increased police presence as a signal that their neighbourhoods 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 programme 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 pre-register 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 2 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 (β = −0.052, 95% CI −0.144 to 0.041, P = 0.271, RI P = 0.431, BF_{10} 0.237) or spillover blocks (β = −0.068, 95% CI −0.146 to 0.010, P = 0.089, RI P = 0.186). We do, however, find some evidence that the programme improved perceptions of safety among business owners; this effect is substantively large, but only weakly statistically significant when using RI P values (β = 0.284, 95% CI 0.056 to 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 (for example extortion), and thus more receptive to military patrols. These explanations are speculative, however, and overall the effects on perceived safety are mixed at best.

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 programme on abuses reported by monitoring (top) and endline survey respondents (bottom). Monitoring survey respondents were asked how many times they had seen or

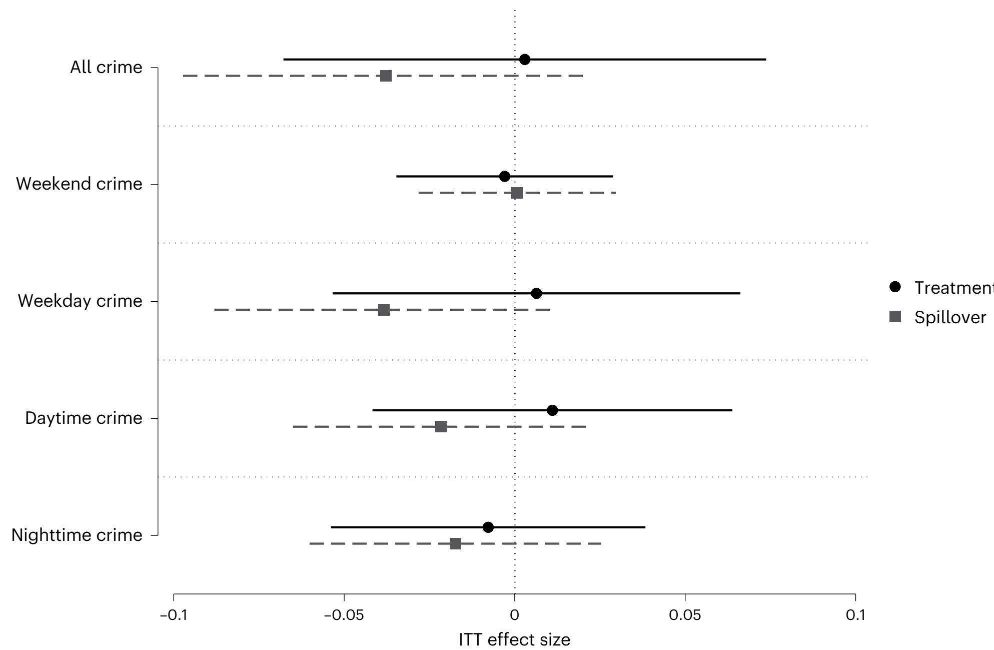

Fig. 2 | Randomized military patrols had little to no effect on crime even while soldiers were physically present on the streets. ITT on crime during the intervention based on administrative data, disaggregated by day and time. The dependent variable in all models is an additive index of murders, armed robberies, thefts, illegal drug sales and illegal possession of weapons. All models include neighbourhood 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% CIs from two-tailed tests for the treatment and spillover ITTs, respectively. No adjustments were made for multiple comparisons. For 'all crime' (N = 1, 167): β = 0.003, 95% CI −0.068 to 0.074, P = 0.934, RI P = 0.957 on treatment blocks; β = −0.038, 95% CI −0.097 to 0.022, P = 0.212, RI P = 0.413 on spillover blocks. For 'weekend crime' (N = 1,167): β = −0.003, 95% CI −0.035 to 0.029, P = 0.857, RI P = 0.890 on treatment blocks; β = 0.001, 95% CI −0.028 to 0.030, P = 0.961, RI P = 0.968 on spillover blocks. For 'weekday crime' (N = 1,167): β = 0.006, 95% CI −0.053 to 0.066, P = 0.834, RI P = 0.901 on treatment blocks; β = −0.038, 95% CI −0.088 to 0.011, P = 0.129, RI P = 0.365 on spillover blocks. For 'daytime crime' (N = 1,167): β = 0.011, 95% CI −0.042 to 0.064, P = 0.681, RI P = 0.786 on treatment blocks; β = −0.022, 95% CI −0.065 to 0.022, P = 0.330, RI P = 0.510 on spillover blocks. For 'nighttime crime' (N = 1,167): β = −0.008, 95% CI −0.054 to 0.038, P = 0.741, RI P = 0.812 on treatment blocks; β = −0.017, 95% CI −0.060 to 0.025, P = 0.424, RI P = 0.514 on spillover blocks.

heard about physical or verbal abuses committed by police officers or soldiers in the past 2 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 (β = 0.010, 95% CI −0.001 to 0.022, P = 0.071, RI P = 0.430) and spillover blocks (β = 0.001, 95% CI −0.005 to 0.007, P = 0.778, RI P = 0.942, BF_{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 ten respondents was surveyed while implementation was ongoing. (They were surveyed on or after 29 November, 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 (β = −0.001, 95% CI −0.009 to 0.006, P = 0.716, RI P = 0.783, BF_{10} 0.075) or spillover blocks (β = 0.002, 95% CI −0.005 to 0.009, P = 0.544, RI P = 0.613, BF_{10} 0.033), with Bayes factors indicative of evidence in favour of the null.

We find more robust evidence of increased abuses by the police. Compared with control blocks, monitoring survey respondents on treatment blocks were 0.037 percentage points more likely to report abuse by police officers during the intervention (β = 0.037, 95% CI 0.016 to 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 (β = 0.016, 95% CI −0.005 to 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 (β = 0.030, 95% CI 0.007 to 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 favour of

https://doi.org/10.1038/s41562-023-01600-1

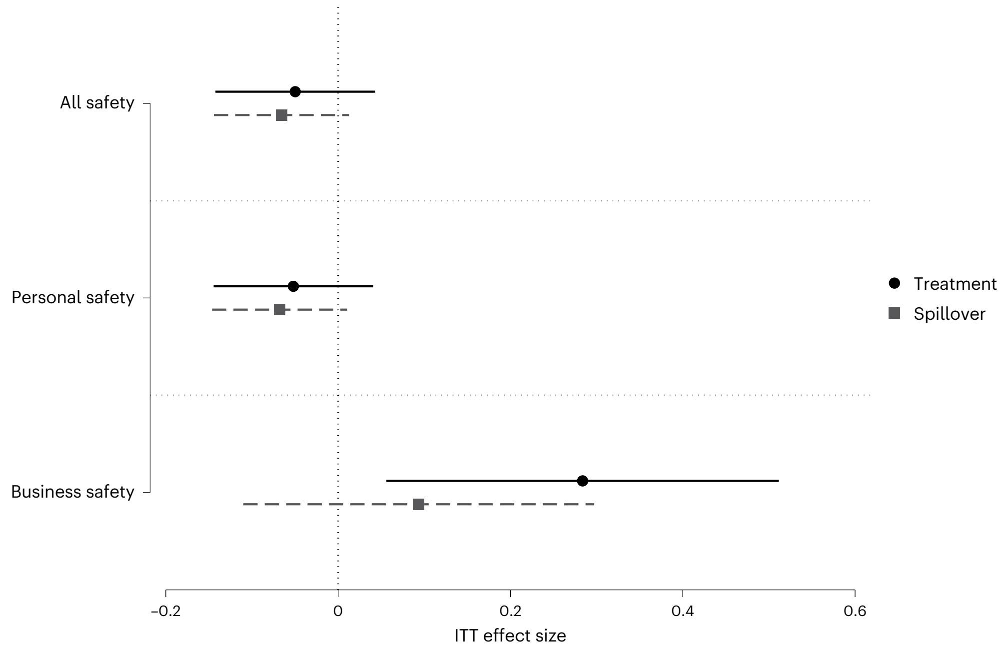

Fig. 3 | Randomized military patrols did not improve perceptions of safety, except among business owners. ITT on perceptions of safety for residents and business owners together ('all safety'), residents only ('personal safety') and business owners only ('business safety') based on endline survey data. 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 2 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 neighbourhood; 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 neighbourhood 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% CIs 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 Supplementary Tables 8 and 9. For 'all safety' (N = 7,707): $\beta = -0.050$, 95% CI −0.142 to 0.043, P = 0.292, RI P = 0.449 on treatment blocks; $\beta = -0.066$, 95% CI −0.144 to 0.013, P = 0.100, RI P = 0.200 on spillover blocks. For 'personal safety' (N = 7,708): $\beta = -0.052$, 95% CI −0.144 to 0.041, P = 0.271, RI P = 0.431 on treatment blocks; $\beta = -0.068$, 95% CI −0.146 to 0.010, P = 0.089, RI P = 0.186 on spillover blocks. For 'business safety' (N = 1,014): $\beta = 0.284$, 95% CI 0.056 to 0.512, P = 0.015, RI P = 0.064 on treatment blocks; $\beta = 0.094$, 95% CI −0.110 to 0.297, P = 0.367, RI P = 0.472 on spillover blocks.

the null ($\beta = 0.011$, 95% CI −0.015 to 0.037, P = 0.417, RI P = 0.541, BF_{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 neighbourhoods 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 who 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 bulletproof vests and travelled 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 ($\beta = 0.105$, 95% CI 0.034 to 0.175, P = 0.004, RI P = 0.064) but no more likely to report police presence ($\beta = 0.032$, 95% CI −0.049 to 0.112, P = 0.439, RI P = 0.611, BF_{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 include 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.

Discussion

What explains the null or even adverse effects of Cali's military policing programme, 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

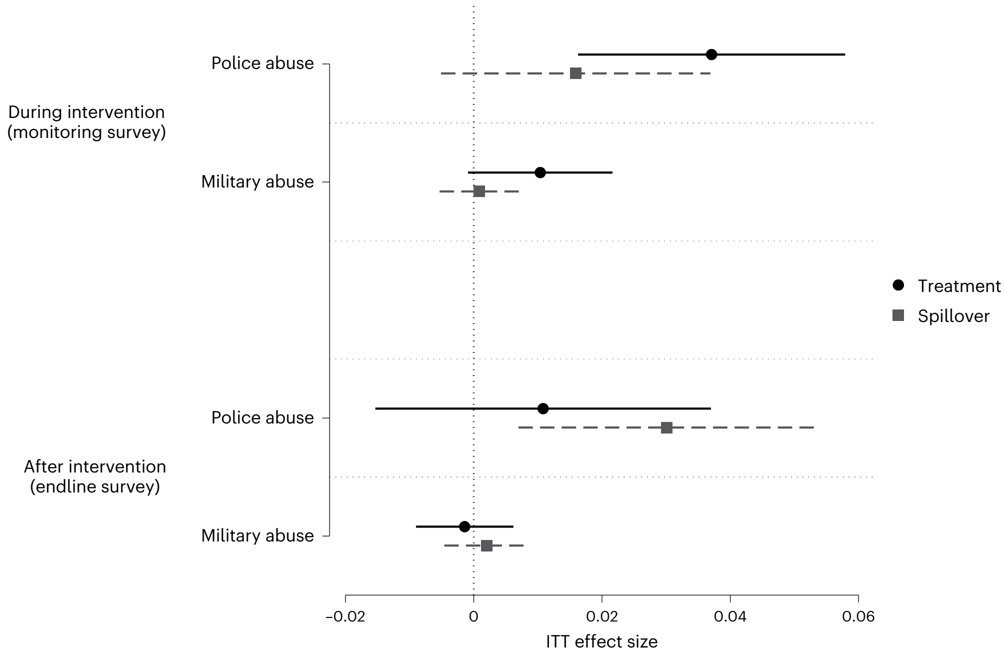

Fig. 4 | Randomized military patrols may have exacerbated human rights abuses, especially by the police. ITT on abuses committed by police officers and soldiers based on monitoring (top) and endline survey data (bottom). The dependent variable in all models is an indicator for any incidents of verbal or physical abuse on the block. All models include neighbourhood 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% CIs 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 Supplementary Tables 10 and 11. For 'police abuse during intervention' (N = 1,970): β = 0.037, 95% CI 0.016 to 0.058, P = 0.001, RI P = 0.103 on treatment blocks; β = 0.016, 95% CI −0.005 to 0.037, P = 0.137, RI P = 0.385 on spillover blocks. For 'military abuse during intervention' (N = 1,970): β = 0.010, 95% CI −0.001 to 0.022, P = 0.071, RI P = 0.430 on treatment blocks; β = 0.001, 95% CI −0.005 to 0.007, P = 0.778, RI P = 0.942 on spillover blocks. For 'police abuse after intervention' (N = 7,908): β = 0.011, 95% CI −0.015 to 0.037, P = 0.417, RI P = 0.541 on treatment blocks; β = 0.030, 95% CI 0.007 to 0.053, P = 0.011, RI P = 0.028 on spillover blocks. For 'military abuse after intervention' (N = 7,908): β = −0.001, 95% CI −0.009 to 0.006, P = 0.716, RI P = 0.783 on treatment blocks; β = 0.002, 95% CI −0.005 to 0.009, P = 0.544, RI P = 0.613 on spillover blocks.

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 artefact 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 (for example, theft and armed robbery) and the types of crimes that respondents reported witnessing in the endline survey (for example, 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, it 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 socio-economic conditions that military policing cannot address. In this case, a more promising deterrence strategy might involve social welfare programmes for citizens or desistance interventions for convicted criminals[48]. The literature on these initiatives is vast and beyond the scope of our study (for a review, see ref. 49). Whatever the explanation, our finding that Plan Fortaleza may have actually exacerbated crime belies one of the key purported benefits of military policing.

Sceptics 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 behaviour, inducing Hawthorne effects. We discuss our approach to minimizing Hawthorne effects in the Methods section; we believe they are unlikely to explain our results. Second, the Plan Fortaleza programme was relatively short, and was confined to weekday nights. As we note in the Methods section, most military policing interventions are similarly short. Still, we cannot know how our results might have changed if the programme 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 microterritories, and gangs are incapable of exerting monopolistic (or even duopolistic) control—an important 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 probably 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, for example, has found that police officers sometimes welcome cameras to discourage 'frivolous, malicious, or unfounded' complaints by citizens (ref. 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.

Methods

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 programme pre-dated 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 neighbourhoods in which the intervention would occur; we randomized only the specific city blocks where soldiers would and would not patrol.

Nonetheless, both the programme 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, 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. Before 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 chequered 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 programmes[55,56].

Finally, although 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. The Ethics Committee did, however, request that we devise a system to ensure that both soldiers and civilian monitors could participate freely in the study. Monitors were hired specifically for the project and were informed about the nature of the study and the intervention, and of any risks and benefits involved. Monitors who did not wish to participate could decline to do so.

Soldiers were similarly informed about the nature of the study and were given the opportunity to ask questions about the methodology. The research team, in coordination with the army and the Security and Justice Secretariat, made any adjustments necessary to ensure that participating soldiers and their commanding officers were comfortable with the design and could execute it faithfully. The soldiers who participated in the study did so as part of their routine authorized activities as members of the armed forces, and were subject to no greater risk than they would have experienced if the evaluation had not occurred. Soldiers who did not wish to join the study could make a request not to participate.

The Plan Fortaleza programme

The Plan Fortaleza programme 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 licences, 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 17:00 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' (neighbourhoods). 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 30 September 2019 and concluded on 18 November 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 min of military patrols roughly every 6 days. In reality, the average time spent patrolling was around 11 min 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 m, with a standard deviation of 248 m.) 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 (ref. 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 (ref. 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 m 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.

Randomization

The 30 neighbourhoods in our sample consist of 1,255 city blocks, with an average of 42 blocks per neighbourhood. We stratified by neighbourhood, then randomized such that approximately 1/6 of all blocks in each neighbourhood 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 modelling 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.

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 9 months before the intervention (1 January 2019) and ending 6 weeks after (31 December 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 17 October and 19 December 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 17 January and 25 February 2020, between 2 and 3 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 2 months, allowing the soldiers to acclimate to their presence, thus further mitigating the risk of Hawthorne effects. Due to the density of these neighbourhoods 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.

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 programme using a weighted least squares (WLS) regression where observations are weighted by the inverse probability of assignment to their realized treatment status. Because the probability of assignment to the spillover and control groups depends on proximity to the nearest treatment block, we cannot calculate inverse probability weights 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 inverse probability weights. In addition, because of the way blocks are distributed across neighbourhoods, 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 WLS regression given by

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

where y_{jk} denotes the outcome for block j in neighbourhood k; t_{jk} denotes assignment to treatment; s_{jk} denotes assignment to spillover; $\mathbf{X_{jk}}$ denotes a vector of block-level covariates; α_k denotes neighbourhood fixed effects; and ϵ_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

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

where y_{ijk} denotes the outcome for respondent i on block j in neighbourhood 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; α_k denotes neighbourhood fixed effects; and ϵ_{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.

Spillover

Our research design allows us to estimate both direct and spillover effects of the Plan Fortaleza programme. 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[58,61]. We follow ref. 40 and report RI P values alongside the more conventional P values and CIs 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[62,63], RI P values provide a useful (and generally more conservative) complement to more conventional test statistics in the presence of fuzzy clustering.

Treatment compliance

Data collected by the civilian monitors suggest that treatment compliance was reasonably high, especially given the difficulty of navigating these neighbourhoods. 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 ($\beta = 0.105$, 95% CI 0.034 to 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 ($\beta = 0.032$, 95% CI −0.049 to 0.112, $P = 0.439$, RI $P = 0.611$, BF_{10} 0.067). This is as expected, and is indicative of treatment compliance.

Deviations from PAP

In our PAP we proposed to test the ITT of the Plan Fortaleza programme 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; here 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 programme 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 ref. 64. 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 behaviour. For compactness we focus in this article 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 behaviour in a separate paper.

Protocol registration

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

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

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

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.

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. *J. Exp. Criminol.* **17**, 473–506 (2021).
2. Flores-Macías, G. & Zarkin, J. The consequences of militarized policing for human rights: evidence from Mexico. *Comp. Polit. Stud.* https://doi.org/10.1177/00104140231168362 (2023).
3. Meliala, A. Police as military: Indonesia's experience. *Policing* **24**, 420–432 (2001).
4. Varona, G. Politics and policing in the philippines: challenges to police reform. *Flinders J. Hist. Polit.* **26**, 102–125 (2010).
5. Montesh, M. & Basdeo, V. The role of the South African National Defence Force in policing. *Sci. Militaria* **40**, 71–94 (2012).
6. Dunlap, C. J. Jr. The police-ization of the military. *J. Polit. Mil. Sociol.* **27**, 397–418 (1999).
7. Braga, A. A. & Weisburd, D. L. *Policing Problem Places: Crime Hot Spots and Effective Prevention* (Oxford Univ. Press, 2010).
8. Braga, A. A., Andresen, M. A. & Lawton, B. The law of crime concentration at places: editors' introduction. *J. Quant. Criminol.* **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 Q.* **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. et al.) 309–316 (Wadsworth, 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. *Perspect. Polit.* **19**, 519–538 (2021).
15. Pion-Berlin, D. A tale of two missions: Mexican military police patrols versus high-value targeted operations. *Armed Forces Soc.* **43**, 53–71 (2017).
16. Zechmeister, E. J. *The Political Culture of Democracy in the Americas, 2014: Democratic Governance across 10 Years of the Americas Barometer* (USAID, 2014).
17. Apel, R. & Nagin, D. S. in *Crime and Public Policy* (eds Wilson, J. Q. & Petersilia, J.) 411–436 (Oxford Univ. Press, 2011).
18. Gibbs, J. P. *Crime, Punishment, and Deterrence* (Elsevier, 1975).
19. Kleiman, M. A. R. *When Brute Force Fails: How to Have Less Crime and Less Punishment* (Princeton Univ. Press, 2010).
20. Nagin, D. S. in *Crime and Justice: A Review of Research* Vol. 47 (ed Tonry, M.) 199–263 (Univ. Chicago Press, 2013).

21. Paternoster, R. The deterrent effect of the perceived certainty and severity of punishment: a review of the evidence and issues. *Justice Q.* **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 Res. Pap. Ser.* **3**, 1–22 (2015).
23. Pion-Berlin, D. & Carreras, M. Armed forces, police and crime-fighting in Latin America. *J. Polit. Lat. Am.* **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. *Comp. Polit.* **51**, 1–20 (2018).
25. Bayley, D. H. *What Works in Policing* (Oxford Univ. Press, 1998).
26. Lessing, B. Logics of violence in criminal war. *J. Confl. Resolut.* **59**, 1486–1516 (2015).
27. Espinosa, V. & Rubin, D. B. Did the military interventions in the Mexican drug war increase violence? *Am. Stat.* **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 Soc.* **36**, 327–350 (2010).
29. Braga, A., Papachristos, A. & Hureau, D. Hot spots policing effects on crime. *Campbell Syst. Rev.* **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. *J. Exp. Criminol.* **15**, 289–311 (2019).
31. Bove, V. & Gavrilova, E. Police officer on the frontline or a soldier? The effect of police militarization on crime. *Am. Econ. J.* **9**, 1–18 (2017).
32. Delehanty, C., Mewhirter, J., Welch, R. & Wilks, J. Militarization and police violence: the case of the 1033 program. *Res. Polit.* **4**, 1–7 (2017).
33. Gunderson, A. et al. Counterevidence of crime-reduction effects from federal grants of military equipment to local police. *Nat. Hum. Behav.* **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. *Am. Econ. J.* **9**, 291–313 (2017).
35. Lowande, K. Police demilitarization and violent crime. *Nat. Hum. Behav.* **5**, 205–211 (2021).
36. Mummolo, J. Militarization fails to enhance police safety or reduce crime but may harm police reputation. *Proc. Natl Acad. Sci. USA* **115**, 9181–9186 (2018).
37. González, Y. M. *Authoritarian Police in Democracy: Contested Security in Latin America* (Cambridge Univ. Press, 2020).
38. Magaloni, B. & Rodriguez, L. Institutionalized police brutality: torture, the militarization of security, and the reform of inquisitorial criminal justice in Mexico. *Am. Polit. Sci. Rev.* **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. *Am. Polit. Sci. Rev.* **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. *J. Eur. Econ. Assoc.* **19**, 2022–2051 (2021).
41. Jeffreys, H. *Theory of Probability* (Oxford Univ. Press, 1961).
42. Lee, M. D. & Wagenmakers, E.-J. *Bayesian Cognitive Modeling: A Practical Course* (Cambridge Univ. Press, 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. *Crim. Justice Policy Rev.* **22**, 350–374 (2011).
45. Rosenbaum, D. P. in *Police Innovation: Contrasting Perspectives* (eds Weisburd, D. & Braga, A. A.) 245–266 (Cambridge Univ. Press, 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. *J. Exp. Criminol.* **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. *J. Exp. Criminol.* **11**, 393–417 (2015).
48. Carr, J. B. & Packham, A. SNAP benefits and crime: evidence from changing disbursement schedules. *Rev. Econ. Stat.* **101**, 310–325 (2019).
49. Doleac, J. L. Encouraging desistance from crime. *SSRN* https://papers.ssrn.com/abstract=3825106 (2020).
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. *Eval. Rev.* **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 Q.* **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. *Criminol. 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* https://www.nber.org/papers/w28393 (2021).
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. *Ann. Appl. Stat.* **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. *J. Exp. Criminol.* **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. Abadie, A., Athey, S., Imbens, G. W. & Wooldridge, J. M. When should you adjust standard errors for clustering?. *Q. J. Econ.* **138**, 1–35 (2023).
62. Gelman, A. A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. *Int. Stat. Rev.* **71**, 369–382 (2003).
63. Keele, L. The statistics of causal inference: a view from political methodology. *Polit. Anal.* **23**, 313–335 (2015).
64. Wang, Y., Samii, C. D., Chang, H. & Aronow, P. M. Design-based inference for spatial experiments under unknown interference. Preprint at *arXiv* https://arxiv.org/abs/2010.13599 (2023).

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 Innovations for Poverty Action Colombia (K. Holloway and S. Jaramillo) for coordination of the project; the Secretaría de Seguridad de Cali (M. A. Arboleda, P. Gracia, A. Sánchez, J. D. Tabares, P. Uribe and A. Villamizar) for logistical oversight; the Third Battalion of the Armed Forces of Colombia (especially O. Arciniegas) for its commitment to the evaluation; and C. Bohm, M. F. Cortés, A. Farfán, S. Flórez, C. Gúzman, S. Hernández and L. M. Mora for terrific research assistance. For helpful feedback we thank P. Aronow, G. Blair, C. Blattman, A. Coppock, E. De Bruin, L. Fergusson, D. Green, A. Moya, M. Offer-Westort, M. Ross, S. Parkinson, S. Tobón, J. Vargas, J. 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
R.B. and M.W. contributed equally to this article.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01600-1.

Correspondence and requests for materials should be addressed to Robert A. Blair or Michael Weintraub.

Peer review information *Nature Human Behaviour* thanks Diego Esparza and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Robert A. Blair and Michael Weintraub

Last updated by author(s): Feb 17, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*
Data analysis	Custom code was created using Stata (version 17) and R (version 4.2.0) to merge different data sources, including GPS data on patrols, administrative crime data, and survey data, all to the block level. All regressions were run in Stata (version 17).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data generated or analysed for this study can be found on Dataverse at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi\%3A10.7910\%2FDVN\%2FWAJ9SR&version=DRAFT.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Sex was self-reported via our surveys. Data on gender was not collected.
Population characteristics	See above.
Recruitment	Enumerators knocked on the doors of households and asked if individuals would like to participate.
Ethics oversight	The Ethics Committee at Universidad de los Andes approved this project (Acta 1073 de 2019 and Acta 1034 de 2019).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We use quantitative experimental methods. In particular, we randomized military patrols at the city block level and then use administrative crime data and survey data to assess the effects of military patrols on a variety of crime and human rights outcomes.
Research sample	Two communes (comunas) in Cali, Colombia. We chose to study Cali because it is among the world's most violent cities, and chose these communes because this is where the government expressed an interest in deploying the policy we evaluated. The 30 neighborhoods in our sample consist of 1,255 city blocks, with an average of 42 blocks per neighborhood. Monitoring survey respondents were 32.1% male and 68.9% female, with an average age of 48.0 years. Endline survey respondents were 33.4% male and 66.6% female, with an average age of 46.4 years. While our sample is not necessarily representative, on average, for our monitoring survey we surveyed three residents and two business owners on each of 416 blocks and for our endline survey on average we surveyed six residents and two business owners per block. See below for more information on sampling strategy.
Sampling strategy	The administrative crime data covers all blocks in both communes. We conducted an original household survey of 2,095 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. We then 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. As our pre-registered pre-analysis plan details, and as we discuss in the manuscript, we conducted power calculations to determine minimum detectable effect sizes. Using our administrative crime data 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.
Data collection	Survey data were collected using tablets. All surveys were conducted in private, where third parties could not hear or see the survey being administered, and respondents were repeatedly reminded that their participation in the survey was voluntary and anonymous, that the survey could be halted at any time, and that they could skip any question they did not want to answer. Researchers were blinded to treatment assignment at the time of data collection.
Timing	Our monitoring data were collected between October 17 and December 19, 2019, while our endline data were collected between January 17 and February 25, 2020.
Data exclusions	There were five blocks (0.3% of the sample) where it was impossible to administer our endline survey due to safety concerns. We drop these blocks from our analysis.
Non-participation	As noted above, there were five blocked (0.3% of the sample) where it was impossible to administer our endline survey due to safety concerns. We drop these blocks from our analysis.
Randomization	Within our two communes we stratified by neighbourhood, 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 treatement 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.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging