State building in the city: The effects of public security and services on crime, violence and state legitimacy in Bogotá*

Christopher Blattman
University of Chicago & NBER

Donald Green
Columbia University

Daniel Ortega
CAF

Santiago Tobón
Universidad de los Andes & IPA†

July 18, 2017

Abstract

Bogotá's government set out to reduce violence and increase state legitimacy by raising state presence on city streets, either doubling police patrol time or delivering clean-up and lighting services. We identified 1,919 high-crime street segments and randomized them to eight months of increased security, municipal services, both, or neither. Interventions at this scale, in a dense network of streets, require us to account for spillovers into control segments. The policy implications also hinge on these spillovers. We show how to design place-based experiments to test for spatial spillovers flexibly, and why randomization inference is necessary to estimate direct and spillover effects. Using administrative data alongside a city-wide survey, we find that increasing state presence reduces insecurity on targeted streets, but has little effect on trust in the state. There is also evidence of increasing returns to state presence, and to targeting the least secure places. But data from all 136,984 city streets suggest that state presence has divergent spillovers. Intensive policing displaces and potentially even increases property crime. But state presence appears to deter violent crime, especially murders and rapes. Municipal services appear to have more positive spillovers, however imprecise. To better understand spillovers, we argue for more experimentation at this scale.

JEL codes: K42, O17, E26, J48, C93
Keywords: crime, violence, police, public services, state building, hot spots, spillovers, field experiment, Colombia

*Acknowledgments: This research would not be possible without the collaboration of the National Police of Colombia and the Mayor's Office of Bogotá, in particular Bogotá's Secretary of Security, Daniel Mejía, who co-conceived the interventions and experiment. Innovations for Poverty Action in Colombia and the Center for the Study of Security and Drugs at Universidad de los Andes coordinated all research activities, survey data was collected by Sistemas Especializados de Información, and for research assistance we thank Juan Carlos Angulo, Marta Carnelli, Daniela Collazos, Eduardo Garcia, Sofia Jaramillo, Richard M. Peck, Patryk Perkowski, Oscar Pocasangre, María Aránzazu Rodríguez Uribe, and Pablo Villar. For comments we thank Thomas Abt, Roseanna Ander, Raquel Bernal, Adriana Camacho, Aaron Chalfin, Marcela Eslava, David Lam, Leopoldo Fergusson, Daniel Mejía, Jan Pierskalla, Tristan Reed, Jacob N. Shapiro, Juan F. Vargas, Dean Yang, and numerous conference and seminar participants. Data collection and analysis were funded by the J-PAL Governance Initiative, 3ie, the Latin American Development Bank (CAF), Fundación Probogotá, Organización Ardila Lülle through its funding for the Center for the Study of Security and Drugs at Universidad de los Andes, the Administrative Department for Science, Technology and Innovation of the National Government of Colombia (COLCIENCIAS), and the J. William Fulbright Program.

†Blattman (corresponding author): University of Chicago, Pearson Institute for the Study and Resolution of Global Conflicts and Harris Public Policy, blattman@uchicago.edu; Green: Columbia University, Political Science, dpg2110@columbia.edu; Ortega: Latin American Development Bank (CAF), dortega@caf.com; Tobón: Universidad de los Andes, Economics, and Innovations for Poverty Action, stobon@poverty-action.org.

1 Introduction

Police and city workers are the everyday face of the state. These street-level bureaucrats are responsible for some of the most basic public goods we expect from government, especially security.[1] The presence of a police patrol, the response to a crime, the picking up of garbage, and the lighting of streets—it is impossible not to notice when they are done poorly.

When crime and violence start to get out of control, these are also the first levers that governments start to pull. Cities step up enforcement, they put more police on the streets, or they light up or clean up the places where crime happens.

In the United States, some of the most popular crime-fighting tactics focus on these levers. More than 90% of police agencies use some form of "hot spot policing," intensifying police time and resources in the highest-crime areas.[2] These tactics typically target units as small as a street segment or even particular corners and addresses. Some cities also change the quality of policing in hot spots, enforcing minor infractions with a "zero tolerance" approach. Another common tactic is to reduce disorder in hot spots through municipal services. Services can make it more difficult to commit crimes, by lighting dark areas or increasing people on the street.[3] Cleaning up streets also signals order and state presence, telling criminals to stay away and telling citizens that the state is looking out for them. Altogether, these policing and services interventions form the foundation of the infamous "broken windows" hypothesis.[4]

This is state building on a very different margin than in weak and conflict-ridden states, but the levers weak states use and the reasons why are not so different. From Afghanistan to Iraq or the Philippines, militaries intensify both security forces and public services in order to increase the monopoly of violence, not to mention improving trust in the government and state legitimacy.[5] In more stable places, like major cities in Colombia, the police and government already have a degree of control and legitimacy on most city streets (though not all). They are increasing state presence on the intensive margin, using police and services to complete the last mile of state building.

This raises a number of questions. How much can further state presence reduce crime and violence? Which levers are most effective, and where? Are there increasing or decreasing returns to more state presence, or more levers? And does targeting state presence at high-crime streets actually reduce overall crime in the city, or does it merely push it around?

We tackle these questions in Bogotá, the capital of Colombia. Two percent of the city's 136,984 streets accounted for all murders and a quarter of all crimes from 2012–15. These "hot spots" received less than 10% of police time and limited public services, which suggests some targeting but not to the same degree of

[1]See Lipsky (1969) for a discussion of police as street-level bureaucrats.

[2]See Weisburd and Telep 2016; Police Executive Research Forum, 2008. It can be as simple as increasing the dosage of policing time, such as the Minneapolis case studied by Sherman and Weisburd (1995), or our case in Bogotá. It also includes heightened levels of traffic enforcement, aggressive enforcement of infractions, and problem-oriented policing. For instance, Groff et al. (2015) study three different forms of intensive policing: proactive "problem-oriented" policing at hot spots, increasing the dosage of policing time, and offender focused interventions at hot spots.

[3]Both intensifying police presence and lighting streets is intended to raise the risk of detection and capture for offenders—a tenet of the economic approach to crime prevention. This is rooted in the rational explanation for crime put forth by Becker (1968), where crime is a gamble and increasing expectations of apprehension and punishment deters people from crime.

[4]See Wilson and Kelling (1982); Apel (2013). Over time, "broken windows policing" has come to mean a sort of intensive, zero tolerance policing. But in principle more visible state presence and physical order should send similar signals to citizens and criminals.

[5]See for instance Berman et al. (2011, 2013); Crost et al. (2014). Besides weakening insurgents directly, intensifying security forces and public services are designed to increase trust in the state. As the state becomes more legitimate, and wins the "hearts and minds" of citizens, the idea is that they will be more likely to inform on offenders, share information, or otherwise collaborate against the insurgents (Berman and Matanock, 2015).

crime concentration.[6] In January 2016, a new city government decided to try increasing state presence in high-crime streets much more. They not only wanted to improve security, they also wanted to raise citizens' trust in police, and improve the legitimacy of the local government.

We worked with the city government and police to identify nearly 2,000 crime hot spots using official geolocated crime data plus station-by-station input from police on the problematic segments in their jurisdictions. Within this sample, the city first doubled police patrol time on 756 street segments (a length of street between two intersections). Then they targeted 201 segments for clean-up and better lighting.

We worked with the city government to randomize assignment to either intensive policing, more municipal services, or both interventions. We selected a large sample of streets with high but nonetheless varying levels of initial crime. Finally, since these interventions could displace crime to nearby streets (and because the benefits of police and services could also diffuse), we designed the experiment to measure spatial spillovers.

The city modeled its interventions on standard U.S. policing practices and evidence. Like Bogotá, crime in large U.S. cities concentrates in a small number of hot spots. Based on several experimental trials, there is a consensus in the U.S. that targeting hot spots with more state presence reduces crime at hot spots.[7] The enthusiasm for intensive policing is based partly on two systematic reviews that argue that the evidence points to a diffusion of benefits to nearby streets (Braga et al., 2012; Weisburd and Telep, 2016).[8]The success of hot spots interventions hinges on whether or not crime is simply pushed around the corner. The aggregate effects on crime are difficult to pinpoint, however, because of the small size of most studies. The median study in the above reviews has less than 30 treated hot spots per treatment, and the largest study has 104. These sample sizes make it difficult to detect effects less than 0.4 or 0.5 standard deviations in size (as we illustrate in Appendix A.1). As a result, the direction of spillovers is not much more reliable than a coin flip, and it is not possible draw firm conclusions about whether crime displaces or benefits diffuse.[9] Bogotá offers an opportunity to test impacts on an unusual scale, enough to identify direct treatment effects of 0.15 standard deviations, and spillovers as small as 0.02 standard deviations.

In a first intervention, the Bogotá Mayor's office reallocated existing police patrols to spend more time on high crime streets. No new police were added in the city. Within their usual patrol area (a quadrant), officers

[6] Authors' calculations based on pre-intervention official crime statistics for Bogotá between January 2012 and September 2015. These data are discussed below.

[7] On increased police presence/resources and crime, Chalfin and McCrary (2017) review the broader evidence, with most papers finding policing decreases crime. Prominent examples are Di Tella and Schargrodsky (2004) and Draca et al. (2011). A Campbell Systematic Review of intensive policing interventions identified 19 experimental or quasi-experimental eligible studies (including 9 experiments). Among 25 tests of the core hypothesis, 20 report improvements in crime. See Braga et al. (2012). With the exception of ongoing evaluations in Medellin (Collazos et al., 2016) and Trinidad and Tobago (Sherman et al., 2014), these evaluations are largely in the U.S.

The evidence on interventions that tackle disorder is more limited, but there are some indications of success. Braga et al. (1999) report significant reductions in crime following a combined treatment of intensive arrests, improvements in the physical environment and provision of social services in a randomized controlled trial at Jersey City. Also, Braga and Bond (2008) conducted an experiment in Lowell, Massachusetts aimed at testing the effect of intensive arrests and environmental interventions. These later included surveillance cameras, lighting and clearance of abandoned buildings. The authors report significant reductions in crime stemming from each of these interventions. There is some evidence that street lighting reduces crime (Farrington and Welsh, 2008). Cassidy et al. (2014) conduct a systematic review of the literature on the effects of urban renewal on youth violence. The authors find five studies suggesting there is week evidence on a causal link (see also Abt and Winship (2016)). One of these studies is in a similar context as the one in Bogotá. Cerdá et al. (2012) study the effects of urban infrastructure developed alongside public cable cars in low income neighborhoods in Medellín. Matching treated neighborhoods with a set of control neighborhoods with similar baseline characteristics, the authors find statistically significant decreases in homicide rates and violence reports.

[8] These studies tend to focus on intensive policing of property and violent crime. Banerjee et al. (2017) see substantial displacement from drunk driving checkpoints in India. Drunk driving prevention is an important but distinct criminal and behavioral phenomenon from property and violent crime.

[9] It is possible that a meta-analysis of the existing studies using the micro data could improve the precision of these spillover estimates, but these micro data are generally not available. Even a weighted average of study effects is difficult to calculate because of variation in statistical procedures used, inconsistent reporting, and other gaps in the information.

were told to double their time on two hot spots from roughly one to two hours a day in multiple visits. This intensive policing lasted from February to October 2016. With an average of 130 segments per quadrant, this had a negligible effect on patrol time on other segments. Patrols simply went about their normal duties, moving about on motorbike and foot, interacting with citizens, and stopping and frisking suspicious people. Shortly after this first intervention was underway, the city decided to tackle social disorder by directing city agencies and contractors to repair lights and clean-up trash, graffiti, and overgrown trees.

We focus on police administrative data on reported crimes to evaluate impacts. Police data are problematic, however, if crime reporting is correlated with treatment status. If police double their time on a street, perhaps citizens are more likely to call them, or the officers are more likely to record a crime. Also, police are not entirely blind to the treatment assignment and could manipulate statistics. Therefore we also conducted a survey of about 24,000 citizens, measuring unreported crimes, perceptions of security, and attitudes to police and the city government. These surveys give us important new outcomes such as citizen's perceived risk, or trust in the government and police. More importantly, the survey data show us that there is little evidence of manipulation or reporting that is correlated with treatment.

We also designed the study to measure spillovers. Treating one hot spot can affect the outcomes of untreated hot spots for several reasons: because criminals may shift their activities to nearby hot spots, because the authorities reallocate resources from untreated to treated segments, or simply because places close to treated segments have to be crossed to deliver the intervention. If criminals move from treated to control hot spots, spillovers pose an identification problem. This could lead us to over or underestimate the true effects of the interventions.[10] But we are also interested in spillovers to neighboring streets outside the experimental sample, or "non-hot spots". Added up, these two kinds of spillovers can tell us whether crimes are deterred or simply pushed around the corner.

We can estimate spatial spillovers with a continuous, monotonic rate of decay (and do so). But since we don't know the pattern of spillovers or the distance over which they travel, however, we prefer—and pre-specified—a more flexible design over multiple possible catchment areas.[11] This approach subdivides the hot spots in the control groups into three donut-like categories, with treated hot spots at the center: 0–250 meters distant; 250–500 meters distant; and >500 meters distant. This approach lets us test for treatment effects on directly treated hot spots as well as separate spillover effects on the hot spots in the 0–250 meter and 250–500 meter regions. We follow a similar approach to estimate spillovers into the non-experimental sample of streets.

Spillovers present other estimation challenges, however. By simulating many experiments, we show that the close proximity of Bogota's high-crime hot spots leads to hard-to-model patterns of clustering as soon as we account for spillovers. The reason is that some streets have very high probabilities of assignment to spillover status, since they are close to other hot spots (e.g. in the city center). This means that, in most randomizations, they are assigned to spillover status rather than a pure control group. This generates clustering patterns among our hot spots that do not correspond to a fixed area, and so we cannot use standard correction procedures (such as clustering standard errors within a geographic unit). As a result, conventional standard errors are much too small. We show that randomization (or permutation) inference provides the exact statistical significance.

Broadly speaking, our results suggest that increasing state presence reduces insecurity, and that there may be increasing returns to state presence. We also find evidence of a heterogeneous response by type of

[10]For example, these hot spots could be the most profitable or least risky places to commit crimes.
[11]For details on all pre-specified aspects of the design see https://www.socialscienceregistry.org/trials/1156.

crime: policing displaces property crime to nearby streets, canceling out many of the benefits, while violent crimes—especially homicides and sexual assaults—decrease.

First, both forms of state presence slightly reduce the number of reported crimes on a segment, as well as people's perceived security risk. Intensive policing and municipal services both reduced insecurity by more than 0.1 standard deviations in directly treated hot spots. These impacts are statistically significant if we ignore spillovers into control hot spots. If we account for this interference between units, however, the direct effects become slightly smaller and significantly less precise. Even the most generous estimates, however, point to modest aggregate effects. We estimate a total of 86 crimes may have been deterred in treated hot spots over the eight months of the interventions. Other specifications suggest it is as few as 8 crimes deterred city-wide.

Second, there is some evidence that intensive policing pushed crime around the corner, especially property crime. We look at the sample of 77,000 non-hot spot segments within 250 meters of the experimental sample of hot spots. Being close to a intensively policed hotspot increases reported crimes, enough that seems to more than cancel out any direct effects of the policing intervention. This estimate is imprecise, but we can fairly confidently rule out a decrease in aggregate property crimes. If anything, these crimes seem to increase.

Third, and more hopefully, the benefits from municipal services may diffuse to neighboring streets. Generally, these positive spillovers are imprecise and must be taken with caution. But if we assume a monotonic and continuous rate of decay of spillover effects, the spillover benefits of municipal services are statistically significant.

Fourth, and perhaps most important, any displacement of crime seems to be solely concentrated in property crime. There is some evidence that the interventions led to a decrease in aggregate violent crimes. For example, we estimate there were 97 fewer homicides and sexual assaults in the city as a result of the intervention (though this decrease is not statistically significant).

Fifth, the effects on crime tended to be greatest in the 75 hot spots that received both intensive policing and municipal services. The difference between getting both and one treatment is not statistically significant, but it points in the direction of increasing returns to state presence on these streets.

Finally, we don't see any evidence that improving state presence increased the average citizen's trust in the state or its perceived legitimacy. If anything, more intensive policing slightly reduced people's opinions of the Mayor's office. It is difficult to say why. It is possible that intense police presence intimidates or upsets some residents, although our qualitative investigations show no evidence of this happening.

Methodologically, this study illustrates the importance of scale in estimating the effects of place-based interventions, and the importance of accounting for interference between treatment and control units. It also shows the importance of using randomization inference as this interference grows, to avoid overstating precision. Even small spillovers (and their estimated precision) can dramatically change our understanding of impacts and any policy conclusions.

For example, many cities and police departments weight serious violent crimes much more than property crimes. Quite reasonably, a city government could embrace the results of these interventions, trading off a modest increase in property crimes for a decrease of roughly 97 homicides and sexual assaults. Alternatively, a city may decide to target only violent crime hot spots and follow a different approach for places where property crime is more prevalent.

We can only speculate why property and violent crimes moved in different directions. The difference between the two is statistically significant even when the treatment effects on each type of crime is not. Perhaps crimes of passion are more easily deterred by police presence and more difficult to "take somewhere else." Meanwhile, property crimes may be more calculated, and career criminals ply their trade elsewhere when state presence rises.

Two results also deserve more investigation and experimentation in future. One is the direct and spillover effects of municipal services. These can be more precisely estimated at larger scale, and by disentangling the effects of lighting versus clean-up. Currently our results suggest a role for both. The second is the effect of more intense treatments, or generalized increases in local policing. It is possible that a general increase in policing would limit how much property crime gets pushed around the corner. This could be tested at a quadrant level.

We also believe these results call for larger-sample policing studies in the U.S. with attention to experimental and non-experimental spillovers. It is unclear whether the effects we see in Bogotá would be replicated in U.S. cities. No study to date has had the power to reject moderate or even large adverse spillovers on property and violent crime, and our result is well within the confidence intervals of the U.S. studies. Large samples are not just needed to estimate the direction of spillovers more precisely. They are also needed for finer distinctions between violent and property crime, or to estimate returns to intensity. They are important not just from a policy perspective but to better understand the nature of crime and displacement.

Our results, if true more generally, add some nuance to a common argument in criminology: that crime and violence are concentrated in a small number of people, places, and behaviors; and that targeted interventions stand the best chance of being effective.[12] Alongside another large-sample study of policing, of drunk driving checkpoints by Banerjee et al. (2017), our evidence reinforces the idea that crime is concentrated, but targeting places may not be effective to the extent that it is simply pushed around the corner in some cases. If place-based interventions simply displace property crime, then targeting people and behaviors could be more impactful to address this kind of criminal behavior.

More broadly still, there are interesting parallels between our results and the historical literature on states, where the most common response to state coercion has been for people to elude the state or run away (Scott, 2014). This is the perennial problem of state building, and the evidence from Bogotá suggests it could hold true even in the last mile of state building.

2 Setting

Bogotá, a city of roughly 8 million residents, is the economic, industrial, and political center of Colombia. In 2015, Bogotá had a GDP per capita of $9,612 at market exchange rates, or about $22,000 according to purchasing power parity or PPP estimates. About 10% of the population was below the national poverty line for metropolitan areas of PPP$6 a day, and about 2% was below the national extreme poverty line for metropolitan areas of PPP$2.50 a day.[13] The poor include a large number of people displaced by a low-intensity civil war, a conflict that ran for a half century until a ceasefire and peace agreement in 2016.

[12]See for example Braga et al. (2012); Abt and Winship (2016); Weisburd and Telep (2016); Weisburd et al. (2017)

[13]Population is from the National Department of Statistics (DANE). It is a projection from the 2005 census. GDP per capita is denominated in 2015 US dollars and taken from the Technical Bulletin on Regional GDP for Bogotá from DANE and corrected for purchasing power parity (PPP). For this correction, we use the Conversion Rates for PPP from the Organisation for Economic Cooperation and Development (OCDE). The equivalent GDP per capita figures for Colombia as a whole are $6,049 or PPP$13,808, and for Latin America it is $8,687 or PPP$15,617 (data for Colombia is from DANE and data for Latin America is from the International Monetary Fund). National poverty thresholds and percentages of the population under these thresholds in Bogotá are taken from the Technical Bulletin on Poverty for 2015 from DANE. The thresholds are corrected for purchasing power parity and set in 2011 US dollars as is the standard definition for the poverty threshold by the World Bank.

2.1 Crime in Bogotá

Crime is one of the most pressing social problems in the city. Murders per 100,000 inhabitants is one of the simplest measures to compare. In 2015, Bogotá's homicide rate was 17.4, and in 2016 it dropped to 15.6. This is considerably lower than some of most violent cities in the world, such as 120 in Caracas (Venezuela), 65 in Cape Town (South Africa), or 64 in Detroit (USA) and Cali (Colombia). It is comparable in crime rates to a U.S. city like Chicago, with 15 murders per 100,000 in 2015, while much greater than the 7 recorded in Los Angeles, or 4 in New York.[14]

While crime in Bogotá has improved, it remains a major public issue. In the 1990s Bogotá was one of the most violent cities in the world, with 81 murders per 100,000 people in 1993. A number of factors are said to have contributed to the improvement, including the decline in civil war, as well as advances in police capacity, gun control policies, restrictions on alcohol consumption, and a major local security push.[15] Despite these gains, as in cities with comparable crime levels like Chicago, crime remains one of the foremost social and political concerns of citizens and the government alike.

Like many cities, crime in Bogotá is highly concentrated. According to official crime statistics, from 2012 to 2015 just 2% of the city's 136,984 street segments accounted for all murders as well as a quarter of all other reported crimes.

The nature of Bogotá's crime varies, from pickpocketing and cell phone theft in busy commercial areas, to burglary of businesses and homes, to drug sales and any resulting violence. But hot spots are distributed around the city. They include wealthy areas where criminals come to mug pedestrians, burgle homes, or steal expensive cars. They include more barren industrial areas with little traffic, where it is easier to sell drugs, or steal, or pimp and prostitute (though pimping is illegal, prostitution is not a crime in Colombia, and prostitutes may be drawn to areas with other illegal activities, or attract them). Hot spots also include popular nightlife areas that lead to bar fights, drug sales and usage, or other disturbances.

Most offenders in Bogotá are individual young people committing petty crimes and assaults. There are some semi-organized youth gangs, and some organized crime in the city, but they do not seem to be responsible for the vast majority of the street crime or violence.[16] One exception is three streets called "the Bronx", that were completely controlled by professional crime networks, which (as described below) received special treatment in this intervention.

2.2 Security policy and policing

Bogotá has relatively moderate to low levels of police compared to large U.S. cities or other large cities in the country and Latin America. There are roughly 18,000 police officers in operational activities in Bogotá, including about 6,200 patrol agents. We estimate the number of active police personnel is about 239 per 10,000 people. The national Colombian average is 350, and virtually all Colombian cities are above Bogotá's police to population ratio. The national ratio in the U.S. was 230 in 2013 but is generally much greater in

[14] U.S. figures come from the FBI Uniform Crime Report and others from the World Atlas. Data for Bogotá was reported by the Mayor's office.

[15] For a detailed review of the policies and programs that improved security conditions in the past decades see Vargas and Garcia (2008).

[16] A study reports that about a third of the homicides in the *localidades* of Kennedy and Ciudad Bolivar in the south of Bogotá are likely to be related to instrumental violence and revenge but concrete ties to organized crime are not clear (Velasquez, 2010). A report by UNODC also documents the presence of organized crime in the south of Bogotá, mostly related first to the presence of left wing militias and then to paramilitaries trying to overthrow the militias' power. These groups then diverted to drug dealing activities that had their distribution centers in Ciudad Bolivar and other places in the south as Corabastos, the main distribution center in Colombia for agricultural products (Beltran et al., 2012).

large cities, including ratios of 413 for New York, 444 for Chicago, and 611 for Washington. The rate in Bogotá is not so different from Los Angeles, which has a rate of 257.[17]

The police freely patrol all city streets, with rare exceptions such as the Bronx. While 2% of streets account for a quarter of the crime, we estimate they received roughly 10% of police patrol time 2012--15.

Police patrols are reasonably well-regarded. The broader police force is not without problems, but our citizen survey (detailed below) suggests that the street patrol officers are regarded as basically competent and generally non-corrupt. When there are complaints, it is usually that residents would like the patrols to spend more and not less time on the street.

In January 2016 a new Mayor came to power, Enrique Peñalosa, and central to his election platform was the reduction of crime, increasing trust in police and government, and ending the influence of organized crime.[18] One of Peñalosa's first actions was to appoint crime economist Daniel Mejía to lead these security efforts. Peñalosa created a new Secretariat of Security for this purpose. Collectively we refer to their organization as the Mayor's office.

The first item on the Mayor's election platform was to tackle crime and violence in the city's 750 highest-crime streets. In his first 100 days, he pledged to dedicate more municipal services and law enforcement to these areas.[19] Privately, the Mayor's office explained that the goal was to not only to reduce crime and violence, but also increase trust in the state and state legitimacy in the eyes of citizens.

The major municipal services that could be targeted at streets included trash collection, tree pruning, graffiti clean-up, and streetlight maintenance. The agencies that coordinate these services report directly to the Mayor's office. Even so, Mayoral control over these agencies has some limits. Much of the work is actually done by private contractors with pre-existing contracts for a particular service in a particular part of the city. In general, the Mayor's office can direct agencies and their contractors to do their job differently, but like any municipal bureaucracy, this can be difficult to monitor and enforce, and it generally needs to fall within existing contracts.

When it comes to the police, the Mayor's office can influence tactics, force allocations, and equipment, but has little say in total force size. City police forces in Colombia are a branch of the National Police and report up to the Minister of Defense, not the Mayor's office. But the city has the power of the purse, giving police officers most of their equipment and paying for various expenses. The Colombian Constitution also gives Mayors authority over all city security matters, and calls on the police to comply with the Mayors' requests and policies.

Changes in force levels are much more expensive, however, and require more cooperation with the national government. In 2016, Mayor Peñalosa requested an increase in police manpower in the city. The national government rejected the request because of a national budget crisis. Instead, to improve security to the 750 hot spots, the Mayor's office focused on increasing the efficiency and quality of the existing police, especially street patrols.

2.3 How does patrolling work?

The quadrant (*cuadrante*) is the most basic patrolling unit. Bogotá has 19 urban police stations, one for each *localidad*, an administrative division used for most municipal services. Stations are divided into CAIs—*Comando de Atención Inmediata*—a small local police base that coordinates patrol agents and takes

[17] See Appendix A.2 for a comparison across cities and data sources.
[18] Peñalosa had previously been Mayor from 1998 to 2001.
[19] See for instance the major national newspaper El Tiempo.

civilian calls. Each CAI has about 10 quadrants. There are 1,051 quadrants in urban Bogotá, and the average quadrant contains about 130 street segments.

Each quadrant has six permanent patrol men and women.[20] They patrol in pairs, on motorbike and foot, in three shifts of eight hours each. In practice, patrols are expected to move more or less continuously in the quadrant throughout their shift, by motorbike. They may patrol a street on motorbike but in general they will dismount regularly to speak to shopkeepers, passersby, and suspicious people.

Patrols carry a handheld computer that allows them to check a person's identification number for outstanding warrants. Stopping and running a person's identification number is a common activity, and patrols have daily quotas to meet. Patrols are expected to regularly stop and frisk any suspicious people, and will seize illegal weapons (usually knives), illegal drugs, and other contraband. Patrols tend to focus these interrogations on young men. If the patrol arrests someone, both patrollers must take the suspect to the main station, where paperwork and other processing can take many hours. This keeps them from meeting their daily performance goals, and so patrols are thought to avoid minor arrests (or perhaps accept bribes in compensation).[21]

The handheld computer also contains a global positioning system (GPS) chip that records the patrol's location roughly every 30 seconds (when operational). The city first piloted and introduced the system in late 2015, under the previous Mayor. The new system lets station commanders view patrol positions in real time and get regular performance statistics. Thus the study period is a period of increased monitoring and measurement of patrol activity

2.4 Hot spot identification and the experimental sample

We worked with the Mayor's office to identify the highest crime street segments, or hot spots. A segment is a length of street between two intersections, and is a common unit of police attention worldwide (Weisburd et al., 2012). We developed an experimental sample of 1,919 hot spots in 779 of the city's 1,051 quadrants. Figure 1 plots these street segments.

We first generated a list of 2,740 segments (about 2% of all streets) using an index of reported crimes. The National Police had provided us with geo-coded official crime statistics for all 136,984 segments in Bogotá from January 2012 to September 2015. We constructed a geo-fence of 40 meters around each segment and assigned a reported crime to that segment whenever it fell within its geo-fence. We ranked segments based on a weighted sum of the crimes of most concern to the Mayor's office: homicides, assaults, robberies, car theft and motorcycle theft.[22] In order not to overextend patrols, we identified no more than four hot spots

[20] About 13% of the police agents in Bogotá are women, as reported by the Mayor's office.

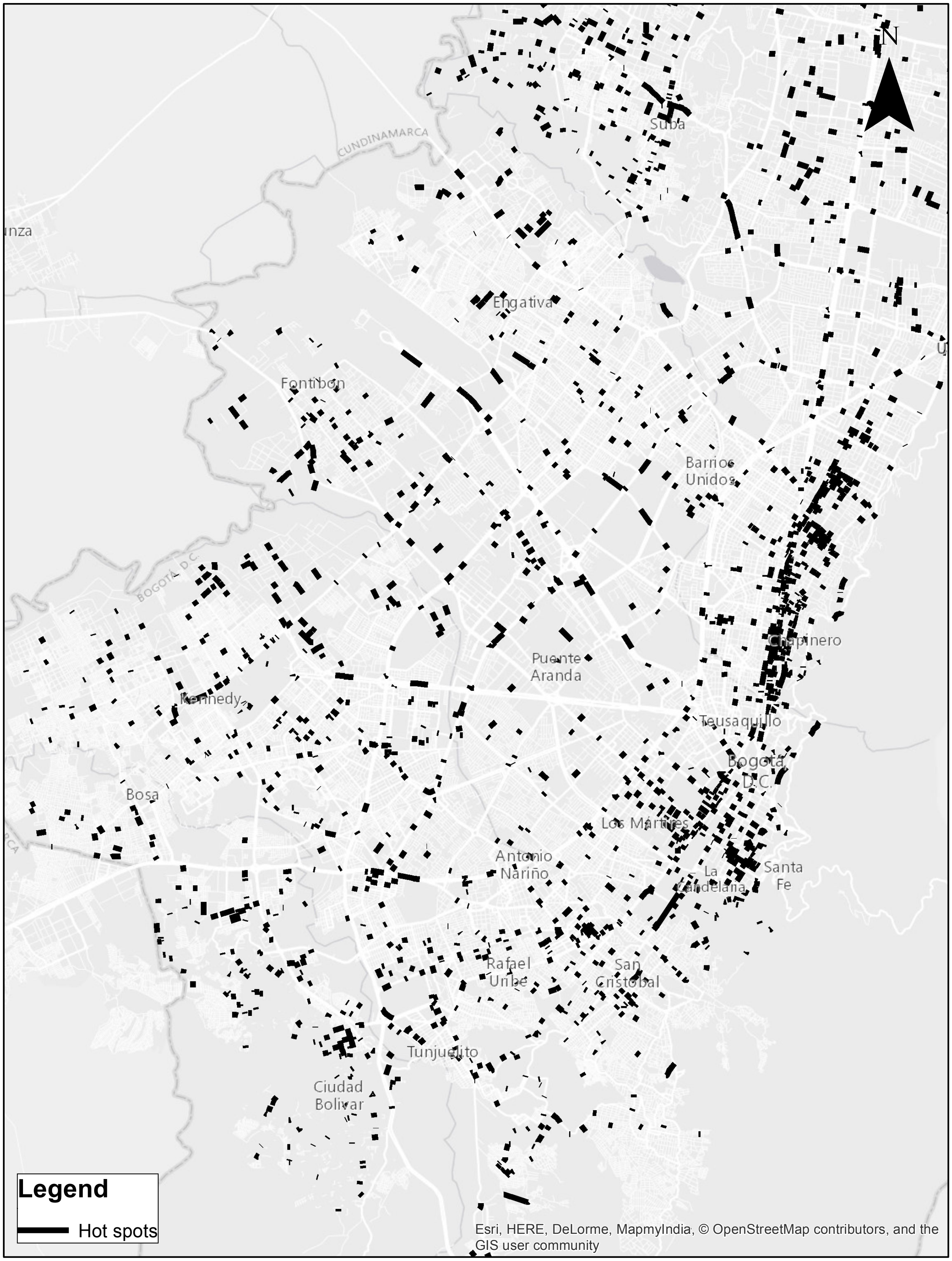
[21] The extent of police corruption in Colombia is largely unknown, although recent scandals point to deep corruption situations. For instance, the first semester of 2016 almost 400 police men and women throughout the country were removed from their jobs and prosecuted because of links to corruption cases. Also, as reported by the major national newspaper El Tiempo, officials estimate that about 5% of almost 190,000 police in Colombia have ongoing investigations related to corruption and misbehavior.

[22] A calculation error meant that 608 segments outside the top 2% were included in this initial sample. These were generally high crime segments, as 90% of those streets were above the 95th percentile of baseline crime, and all were above the 75th percentile. In retrospect, this error proved useful since it gave us more variation in baseline crime levels, which we use to study treatment heterogeneity.

If there is a crime within two or more geo-fences, we assigned the crime to the closest segment using linear distances. Thus if a crime occurred in a public park, it would be assigned to the nearest segment. There are some missing data, especially in the first two years of the data, when about a quarter of reported crimes could not be geo-coded because of deficiencies in the address data. From 2014 onwards, the crime data come with a geographically coordinate, but in some cases these coordinates do not fall within any 40 meter fence and were therefore not assigned to a segment. It was also possible for crime locations to be mis-recorded by the police or citizen

We based crime weights on the average prison sentence according to Colombian law, which proxy for the social costs of crime. For the aggregate crime index, weights are: 0.300 for homicides, 0.112 for assaults, 0.116 for theft from person and 0.221 for car and motorcycle theft. The weight for homicides was cut by half in order to avoid every segment with one homicide in the

Figure 1: Map of hot spots



Notes: Hot spot street segments, in red, are the 1,919 streets included in our experimental sample, as described in Section 2.4 below.

per quadrant (so that no more than two would be assigned to treatment).[23]

Official crime data omitted many forms of petty crime and disorder, plus unobserved risk factors, however, and so we were not satisfied with a mechanical ranking of segments.[24] As a result, we sat with the commanders and some lower-ranking patrol agents from each of the 19 police stations to verify the hot spots. The police eliminated about a third of the hot spots, adding others in their stead, leaving 1,919 segments that account for 21% of the city's reported crimes.[25]

Table 1 reports summary statistics for these 1,919 segments, including crime statistics, other street characteristics, and city service data.[26] In October 2016, the police updated all 2012–16 crime data with more accurate GPS coordinates and cleaned data.[27] We report both the original crime data (used to identify hot spots) and the updated data.

The average hot spot had between 0 and 82 crimes reported in the previous four years (461 with the updated data), with an average of 5 crimes.[28] More than half were property crimes, but violent crimes such as murders and assaults were second in importance. 95% of hot spots had relatively low levels of physical disorder such as garbage.

3 Interventions

3.1 Intensive policing

In February 2016, the new Mayor announced that 750 hot spots would begin receiving intensive policing. The intervention was the subject of intense media coverage and public debate. The government, however, did not publicize the eligible high-crime streets, the existence of an experimental design, or which specific streets were being targeted. Intensive policing began on February 9, 2016 and ended on October 14, 2016, with no changes in the targeted segments over the period.[29]

Intensive policing generally meant an increase in police patrol time on the street by about two-thirds or more. As we will see below, we estimate that during the intervention control streets received 86 minutes of

past four years to become a hot spot. For the violent crime index, weights are: 0.439 for homicides and 0.170 for assaults. For the property crime index weights are: 0.345 for car theft from person and 0.655 for car and motorcycle theft. At the Mayor's office direction, we did not use data on family violence, sexual assault, shoplifting, threats, and other lower frequency crimes to determine hot spots. A focus on homicides, vehicle theft, and robbery is also consistent with evidence from U.S. cities that these crimes respond most elastically to increased police presence (Chalfin and McCrary, 017b).

[23]The police believed more than two segments per quadrant, or more than 770 segments in total, would be too cumbersome for patrols and distract them from the full quadrant. They based these guidelines on U.S. evidence from Koper (1995) and Telep et al. (2014), who have argued that there were decreasing returns on crime control after 15 minutes of police presence.

[24]For instance, while homicides were generally recorded by the police, for any other crime to be included in the database, victims had to travel to one of 19 police stations, file a formal report, have the report accepted, and include relevant details such as location (by tapping on a touch screen, for example). Calls or informal reports to police do not show up in official statistics, and police or CAI stations cannot record crimes they observe unless it represents an operational action (illegal drug or gun seizures, for instance). Indeed, our endline survey (discussed below) suggests that official statistics record only about a fifth of all crimes.

[25]Notably, most of the streets added by police had no reported crimes in the 2012–15 police database. The police nonetheless perceived them to be hot spots because they were known as areas of unreported crime such as pickpocketing, drug sales, or muggings. In eliminating streets, the police said that they dropped segments that they suspected had erroneous crime levels because of their location. For instance, streets close to a police or CAI station, a bus station, or a hospital might have too many crimes in the administrative data, because they were incorrectly designated as the crime site.

[26]Appendix B.2 reports baseline summary statistics on the full sample and the experimental sample.

[27]Some crimes moved to nearby segments, and the correlation between the old and new data is 0.35 at the segment level and 0.86 at the quadrant level. These corrections were unrelated to this study.

[28]Quadrants with at least one hot spot had an average of 3.5 reported crimes per segment across the whole quadrant, while the average quadrant in the whole city reported 1.5 crimes.

[29]The Mayor's office initially planned to run this intensive policing intervention for at least 4 to 6 months. They decided to continue the intervention for nearly 8 months in part to permit the research team enough time to raise funds for, design, and conduct a large-scale survey of citizens to evaluate the intervention.

Table 1: Descriptive statistics for the experimental sample (N=1,919) and tests of balance (treatment versus all control streets, including potential spillover streets)

	Summary statistics				WLS test of balance			
					Intensive policing		Municipal services	
	Mean	Std. Dev.	Min.	Max.	Coeff.	p-val	Coeff.	p-val
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
# of reported crimes on street, 2012-15 (original)	4.53	5.72	0	82	-0.17	0.62	-0.13	0.70
# of violent crimes	1.88	2.94	0	56	-0.18	0.21	-0.05	0.75
# of property crimes	2.66	3.97	0	50	0.02	0.95	-0.08	0.76
# of reported crimes on street, 2012-15 (updated 10/2016)	5.18	18.24	0	461	-0.21	0.86	-0.36	0.79
# of violent crimes	1.40	5.38	0	78	0.39	0.38	0.22	0.68
# of property crimes	3.78	14.09	0	407	-0.60	0.45	-0.58	0.52
Average # of reported crimes per segment in quadrant, 2012-15	3.56	5.13	0	61	-0.30	0.50	0.38	0.49
Daily average patrolling time (11/2015 – 01/2016), minutes	38.03	70.27	1	1029	-1.77	0.73	3.42	0.57
Rating of baseline disorder (0–5, + more disorder)	1.18	0.74	0	5	-0.05	0.31	0.35	0.00
Eligible for municipal services	0.86	0.35	0	1	-0.02	0.27	0.22	0.00
Meters from police station or CAI	551.37	351.46	6	2805	-26.18	0.26	-11.95	0.64
Zoned for industry/commerce	0.38	0.49	0	1	-0.09	0.01	0.05	0.16
Zoned for service sector	0.13	0.34	0	1	0.02	0.33	0.03	0.25
High income street segment	0.07	0.25	0	1	0.00	0.79	-0.01	0.54
Medium income street segment	0.55	0.50	0	1	-0.06	0.06	0.00	0.98
# of street segments in quadrant	127.21	86.99	2	672	2.05	0.71	-3.04	0.57
# of hot spot segments identified in quadrant	3.67	2.68	1	14	-0.30	0.08	-0.16	0.31
# of segments treated with intensive policing in quadrant	1.15	0.95	0	3	1.35	0.00	-0.01	0.91
# of segments treated with municipal services in quadrant	0.66	0.69	0	3	-0.08	0.06	0.91	0.00
Intensive policing assignment: Treated	0.48	0.50	0	1	1.00	-	0.00	-
Intensive policing assignment: Proximal spillover	0.29	0.46	0	1	-0.56	0.00	0.01	0.83
Intensive policing assignment: Distant spillover	0.14	0.35	0	1	-0.28	0.00	0.00	0.96
Intensive policing assignment: Pure control	0.09	0.28	0	1	-0.17	0.00	-0.01	0.72
Municipal services assignment: Treated	0.41	0.49	0	1	0.00	-	1.00	-
Municipal services assignment: Proximal spillover	0.19	0.39	0	1	0.05	0.01	-0.31	0.00
Municipal services assignment: Distant spillover	0.17	0.37	0	1	-0.01	0.71	-0.28	0.00
Municipal services assignment: Pure control	0.23	0.42	0	1	-0.04	0.03	-0.40	0.00

Notes: Columns 1–4 display the summary statistics for our sample of 1,919 hotspots, weighted by the probability of being in the observed experimental condition. In columns 5–8, we perform a balance test for treated vs all control units using weighted least squares.

patrol time on average, with treated streets receiving an additional 65 minutes—an 75% increase.[30]

Commanders told patrols to visit treatment hot spots at least 6 times per day for roughly 15 minutes each visit. In hot spots near bars and night clubs, half the visits were to be in the evening. Otherwise they were to do the majority of visits during the day. The police stations and controls generally did not know what hot spots were in the control group, but in principle they could make reliable guesses.

As discussed above, police generally perform a number of standard activities on patrol, and their instructions were to continue to simply perform these normal duties in the target hot spot: running criminal record checks; stopping, questioning, and frisking suspicious persons; door-to-door visits to the community; conducting arrests or drug seizures; and so forth. Intensive policing meant that they performed these duties more visibly and frequently on the targeted segments.

The only exception was in the three streets known as the Bronx. Early in our intervention period, the police and city invaded and cleared the three streets. This was a much more intensive, one-time intervention. Two of the three streets happened to be assigned to treatment and one had been assigned to the control group. We discuss the invasion in more detail below.

The government could not increase the total number of police or patrols in the city, and so an extra 65 minutes of policing time on each of the two targeted hot spots implied that patrol time fell on other segments in the quadrant by roughly one minute on average (given that there were 130 segments in each quadrant on average).

3.2 Municipal services

The second intervention consisted of instructing municipal contractors to deliver their cleanup and maintenance services to selected hot spots. This intervention did not receive as much publicity or attention as the intensive policing. Two different offices within the Mayoral Administration are in charge of conducting municipal services. A first office carries out activities regarding street lights and a second office is in charge of all other activities. Both offices contract private companies to provide each type of municipal services, including repairing street lights, tree pruning, cleaning non-artistic graffiti, and collecting garbage.

Contractors were expected to perform their usual duties, but the Mayoral offices gave the contractors lists of segments where they were asked to first diagnose the issues (i.e. visit to see which activities were needed) and then ensure the appropriate services were delivered, reporting back to the Mayoral office. Municipal services were supposed to begin on April 11, 2016 and continue until the end of the intensive policing intervention.

3.3 How do the Bogotá interventions compare to other hot spots interventions?

In comparison to the U.S. interventions previously studied, we suspect that the Bogotá intervention is broadly similar in style and approach, but as a consequence of focusing on many more streets, the intensity of treatment on treated hot spots in Bogotá is lower than in comparable U.S. experiments.

[30] Before the intervention, 1–2 weeks of GPS data suggested that hot spots received at least 38 minutes of patrol time per day. The police expected that patrolling time would rise with the new monitoring technology, but it is highly doubtful that actual patrolling rose from 38 to 86 minutes. Rather, the 38 minutes was probably an gross understatement of average patrolling time per hot spot. The police did not have much data on pre-intervention patrol times, since the handheld computers with GPS chips were not handed out to patrols until the end of 2015. They piloted the devices from November 2015 through January 2016, during which the streets in our experimental sample of 1,919 hot spots received roughly 38 minutes of patrolling time per day on average and non-hot spots received about 15 minutes of patrolling time per day. These are lower bounds of the effective patrolling time, because these data were subject to limitations as the devices used to measure patrolling time were still being tested. See appendix B.8.

Intensive policing can take many forms. The India drunk driving study is exceptional for evaluating a specific behavior rather than violent and property crimes. Some hot spots interventions resemble invasions of high crime areas and involve mass arrests, like Bogotá's approach to the Bronx. More commonly, it involves more police time on the streets, focusing in violent and property crime, as in the other 1,916 high-crime streets of Bogotá.

But even here approaches vary. More policing can come from more officers in total or a simple reallocation of existing police. The intensity can vary from extra minutes to many hours. And the nature of policing can vary significantly. Some interventions take a zero tolerance approach, enforcing the most minor infractions. Others focus on "problem-oriented policing," where officers try to proactively address underlying problems, or work with communities on solutions, rather than reactively arresting people only. Thus our results are not comparable to zero tolerance approaches to policing, problem-oriented tactics, or dramatic increases in police time or even invasions.

It is hard to compare the Bogotá intervention to the U.S. ones, as many of the studies (reported in Appendix A) do not discuss control group policing or the specific intensity of treatment. In the Minneapolis Hot Spots experiment, for example, intensities were similar to Bogotá, with patrol times about 2.5 larger in treated hot spots relative to controls (Sherman and Weisburd, 1995). Two experiments seem to report more intensive treatments. In the Philadelphia policing tactics experiment, for instance, patrol times were about eight hours per day in treated hot spots (Groff et al., 2015),[31] and in the intensive patrolling intervention of the Jacksonville policing experiment patrols were 53 hours per week in treated hot spots (Taylor et al., 2011).[32] Though no data were provided for control hot spots in any of these studies, we presume the treatment was more intense than in Bogotá.[33]

4 Data

Bogotá has rich administrative data, but these sources have no information on certain outcomes of interest (such as state legitimacy), and the crime data were of questionable completeness (including a danger of measurement error correlated with treatment status). As a result, we complement the administrative data with primary data collection. In the end we draw on six main sources of data prior to, during, and at the conclusion of the interventions.

1. **Administrative data on police and municipal services compliance.** The police shared the full database of GPS patrol locations for all 136,984 streets, 2015–16.[34] City agencies also shared reports on their diagnosis of each street and compliance with treatment for all streets assigned to the municipal services treatment.

2. **Crime and police operations data.** Police also shared the full database of reported crimes and operational results 2012–16, geolocated to all 136,984 streets.[35] Many U.S. studies also use emergency

[31] The Philadelphia policing tactics experiment evaluates three different tactics: problem-oriented policing, foot patrols and offender focused interventions. The patrol times we report here are for the foot patrols intervention.

[32] The Jacksonville policing experiment evaluates two interventions: problem-oriented policing and intensive patrolling..

[33] Appendix **??** describes these other studies in detail.

[34] Not all handheld computers were functional at all times, and at times over 2016 the system went offline for a few days to a few weeks, and so we use data only during those periods when the system was generally operational in a given police station—on average XX of the YY weeks of the intervention.

[35] Prior to the intervention, we received the 2012-2015 data on crimes prioritized by the Secretary of Security to conduct the hot spots selection: homicides, assaults, robberies, and car and motorbike theft. Importantly, 77% of the crimes had the exact coordinates and the remaining 23% had the address, which we geolocated using an ArcGis add-in tool that matched the address

call data since they are less prone to influence and manipulation, but these data were not available for this study.[36] Since crimes are reported in central stations they are likely to be underreported but not necessarily correlated with treatment, as patrols do not make reports.

3. **Survey of Bogotá residents on crime experiences and government attitudes.** In October 2016 we surveyed 24,000 citizens on 2,399 segments—the 1,919 in the experimental sample, plus a representative sample of 480 segments outside the experimental sample. We interviewed 10 people per segment and average responses over each segment.[37] The 15-minute, anonymous survey used a convenience sample of adults available and willing to speak during the half day the interview team was present on the street segment. In addition to demographics, the survey collected several outcomes: (i) perceptions of security risks on the segment; (ii) perceived incidence of crimes on the segment; (iii) crimes personally experienced in general and on the segment; (iv) crime reporting (and reasons for or against); (v) various measures of trust in and perceived legitimacy of the police and the Mayor's office; and (vi) additional data on police corruption, presence of street gangs and extortion.

   To see why survey-based crime reports are important, Figure 2 illustrates the often large differential between experienced and reported crimes. We asked interviewees whether or not they had experienced a crime since the beginning of the year (on any segment in the city) and, if so, whether they had attempted to report it, and if they were successful. Homicides are reported by police if individuals did not report them, so administrative data probably capture most murders. But for other crimes, successful reporting rates range from 59% for motorbike theft and 53% for car theft (where a report is sometimes necessary for insurance) to as little as 23% for a personal robbery or 13% for the theft of car parts. Across all crimes except homicide, about 27% of the people say they reported the crime, and an additional 9% of people say they attempted to report the crime but were unsuccessful,. If these rates were representative of most of Bogota (although this is not a representative sample of streets) then these figures imply that roughly 73% of all crimes in Bogota were not reported in 2016.

4. **Survey of street disorder.** As discussed in section 5.2, to measure levels of street disorder, we sent enumerators to street segments in the sample, to take photographs and rate the presence of graffiti, garbage, boarded-up buildings, and run-down buildings on a 0–5 scale.[38]

5. **Administrative data on pre-treatment street characteristics.** The city also shared data on pre-treatment street characteristics: urban density, income level (high, medium, low), economic use

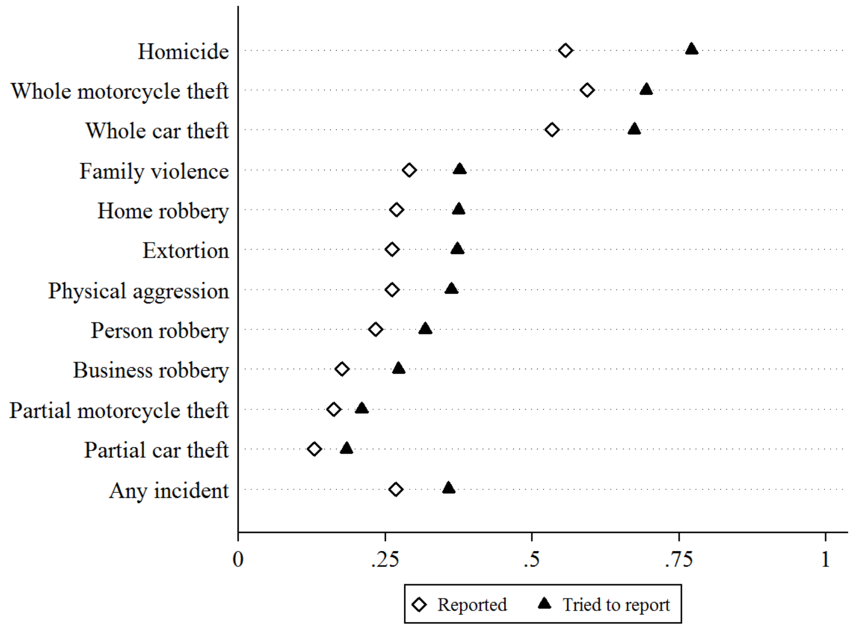
to its approximate coordinates. The success rate of the geolocation process for addresses was 71%, hence about 93% of the reported crimes between 2012 and 2015 were matched to the corresponding segments. We also received all data on arrests; gun, drugs and merchandise seizures; and stolen cars and motorbikes recovered. In October 2016 the police provided updated data that corrected for geolocation problems (thus retrospectively changing pre-intervention data). With the new information we also received data on reported cases of burglary, shoplifting, sexual assaults, family violence, threats, extortion and kidnapping.

[36] Due to continuing technical problems, since the end of 2015 the emergency call system in Bogotá was undergoing a general upgrade to a more reliable technology. Hence, beyond its availability, the integrity of the emergency call data was questionable. See for instance a report from El Espectador.

[37] The exceptions are the three Bronx segments post-invasion where there were no residents to answer questions, and one segment with 11 interviews. Thus, the Bronx segments are not included in any analysis with survey questions. For the segment with 11 surveys, our outcomes are segment averages so we don't need to reweight for this exception. Respondent characteristics and response rates are balanced by treatment status (see appendix B.4 for more details).

[38] We visited 1,534 of a total of 1,919 scheduled streets in March (three months before the municipal services intervention began) in order to narrow down the number of eligible hot spots. We did not collect data in the remaining 385 streets because of security concerns from the enumerators. As we discuss in section 5.2, 1,459 were eligible for the municipal services interventions and 414 of them were assigned to treatment. Those streets were split in two batches of 201 and 213 streets respectively in order to randomize timing, but only the first batch was effectively treated. Then, in order to assess the levels of compliance, we sent enumerators to the 414 streets in the first and second batches in June (one to two weeks after municipal services started to be delivered) and December (two months after the end of the intervention). Again, because of security concerns of the enumerators, we visited 409 in June and 410 in December.

Figure 2: Proportion of crime reported, by crime (survey-based)



Notes: The figure includes data on all street segments surveyed. Each observation is a survey. The white diamonds denote the proportion of people that effectively reported a crime out of all victims. The black triangles denote the proportion of people that tried to report a crime out of all victims.

(housing, services, industry), presence of public surveillance cameras, and distance to the closest police station, commercial area, school, religious center, health center, transport station, or other public services as justice.

6. **Qualitative observation and interviews.** We began with informal qualitative interviews with dozens of police officers and citizens about their experiences with the intervention and police tactics in general. We also hired observers to discreetly visit 100 streets in the experimental sample for a day and passively observe (and document) police behavior. They also interviewed citizens in each segment about police behavior and attitudes.

To simplify our analysis and deal with the problem of multiple comparisons, our pre-analysis plan distinguished primary from secondary outcomes, and pooled like measures into summary indices to reduce the number of hypotheses tested (following Kling et al. 2007).

Our primary outcomes are the effects of the interventions on two measures of insecurity: perceived risk and crime incidence. Table 2 reports summary statistics on a standardized index of each outcome for each of the 4×5 experimental conditions, using inverse probability weights for assignment into each of the treatment conditions. We discuss secondary outcomes, particularly the perceived legitimacy of the police and local government, in Section 6.5 below.

Perceived risk of crime and violence on the segment Our citizen survey asked respondents to rate perceived risk on a 4-point scale from "very unsafe" to "very safe" in five situations, such as: for a young woman to walk alone after dark on this street; for someone to talk on their smartphone on this street; for a young man to walk alone after dark on this street; and simply the perceived risk of crime "during the day" and "at dusk". We construct a index of perceived risk that takes the average across all respondents in the segment. All indexes in the paper are standardized to have mean zero and unit standard deviation.

Crime incidence on the segment We construct a single standardized index of crime that equally weight the survey and administrative data. The two components include: (i) survey respondents' opinion of the incidence of crime on that segment, as well as personal victimization on that segment since the beginning of the year; and, (ii) the total number of crime incidents on that segment reported in the administrative crime data since the beginning of the intervention. We can subdivide all measures into property and violent crimes, although our primary measure pools all crimes into one index. This index weights survey measures twice as much as administrative data.

The survey measured perceived incidence and personal victimization by walking respondents through a list of 11 criminal activities. After finding out whether any of these activities have happened on the street since the beginning of the year, we asked respondents about each crime to establish perceived frequency (ranging from "everyday" to "never" on a 0-6 scale), and whether it happened to the respondent him or herself on that segment. Our results will show results for the three individual components in order to give a sense of the absolute impacts and differences between survey and administrative data.

5 Methodology

The size and direction of spillovers drive the policy implications of place-based anti-crime programs. If crime is simply pushed around the corner, or to nearby hot spots, then the effect on total crime is much less than the direct treatment effects lead us to believe. Many U.S. studies argue that the opposite is true—the benefits of intensive policing diffuse to nearby streets (Braga et al., 2012; Weisburd and Telep, 2016). Given the huge number of nearby streets, even very small spillovers can be consequential. The case for intensive policing hinges on the answer.

Failing to account for spillovers could also bias our estimates of direct treatment effects. If control hot spots are close enough to treated hot spots to experience displacement or diffusion, then spillovers violate the standard assumption of "no interference between units." Previous studies have generally ignored the possibility of interference between treatment and control hot spots, and focused instead on the spillovers into nearby non-hot spots. This is reasonable if the hot spots are widely spread and the catchment areas (or spillover regions) do not overlap. Previous studies have mostly used a catchment area of two blocks or about 150 meters. We generally don't know if there was overlap in these areas, or if spillovers traveled further than two blocks. But the samples in these previous studies are small enough that it may not matter much for the estimates. We do not know. But we certainly cannot ignore this interference as we scale up to hundreds of treated hot spots in a particular city.

The same would be true of any intensive intervention in a spatial or social network. This is a growing source of experimental work. Interference between units is relatively simple to deal with if there are no spillovers between clusters or strata. But if spillovers can spread within the full spatial or social network, then estimation becomes more complicated. We illustrate how to approach these challenges through the design of the experiment and randomization inference.

Table 2: Summary statistics for the primary security outcomes, all experimental conditions

			Municipal services assignment				
			Treated	<250m	250-500m	>500m	Ineligible
			(1)	(2)	(3)	(4)	(5)
A: Perceived risk (z-score)							
Intensive policing assignment	Treated	Mean	-0.073	0.430	0.138	-0.013	-0.373
		SD	0.876	1.017	0.864	0.943	0.934
		N	75	154	150	201	174
	<250m	Mean	0.168	0.335	0.223	0.160	-0.124
		SD	1.061	1.005	0.859	1.369	1.013
		N	74	213	130	125	162
	250-500m	Mean	-0.105	0.291	0.057	0.256	-0.337
		SD	1.042	0.883	0.938	0.942	0.974
		N	32	32	75	80	75
	>500m	Mean	-0.174	0.320	0.124	-0.218	-0.651
		SD	0.914	1.078	1.042	0.912	0.994
		N	20	14	13	68	49
B: Crime incidence (z-score)							
Intensive policing assignment	Treated	Mean	-0.079	0.379	-0.056	-0.047	-0.179
		SD	0.808	1.010	0.790	0.868	0.877
		N	75	154	150	201	174
	<250m	Mean	0.157	0.425	0.139	0.169	0.248
		SD	1.032	1.056	0.849	1.769	1.230
		N	74	213	130	125	162
	250-500m	Mean	-0.143	0.207	-0.053	0.096	-0.105
		SD	0.825	1.024	0.889	0.921	0.874
		N	32	32	75	80	75
	>500m	Mean	-0.215	0.361	-0.147	-0.325	-0.419
		SD	1.092	1.297	1.024	0.745	0.862
		N	20	14	13	68	49

Notes: We report weighted means for each experimental condition, where weights are the inverse of the probability of falling in the corresponding treatment condition. We estimate that probability with repeated simulations of the randomization procedure. The ineligible condition in Column 5 reflects those streets that did not exhibit any disorder at baseline. Technically there are 3×4 ineligible conditions for each dependent variable, one for each relative distance from municipal services treated streets, but we pool those columns here for simplicity.

5.1 Design-based approach

We did not know the potential range of spatial spillovers, and so we pre-specified a flexible design that looked for spillovers in radii of 250 and 500 meters around treated streets.[39] There are many other ways to model spillovers, and we will also show estimates that model a continuous rate of decay, as well as different radii. In our preferred approach, each segment in the experimental sample is assigned to either the treatment or control condition, and we can partition control segments into one of three experimental conditions according to their distance from the treated segment: proximate spillover (<250 meters distant from a treated segment), distant spillover (250–500 meters), and "pure control" status (>500 meters).

Figure 3 illustrates this classification scheme. In this example, the hot spot segment at the center of the two radii has been assigned to the intensive policing treatment with the randomization procedure described below. For simplicity we ignore municipal services in this example. The figure highlights other hot spots in the experimental sample that are nearby, and their treatment condition as a result of their distance from the treatment hot spot.[40]

One virtue of this approach is simplicity and transparency. In essence, all treatment effects estimates are simply differences in the means of the experimental conditions in Table 2. We can also use this design to assess spillover effects outside the experimental sample, on non-hot spot segments, to get high-powered spillover estimates. We opt for regression-based estimates to control for possible confounders, as described below, but these preserve the spirit of the mean differences approach.

It is important to note that this approach ignores the possibility of spillovers beyond 500 meters, as well as non-spatial spillovers. Some crime is undoubtedly displaced in non-Euclidean ways (e.g. to possibly distant hot spots where the benefits of crime are high and the risk of detection is low).[41]

5.2 Randomization procedures

We used a two-stage randomization procedure to maximize the spread between hot spots assigned to each treatment. This helped ensure that as many segments as possible had a high probability of assignment to the pure control and distant spillover conditions. We first blocked our sample by the 19 police stations, then randomized hot spots to intensive policing in two stages: first assigning quadrants to treatment or control, then assigning hot spots within treatment quadrants. This approach also helped us to meet the office of the Mayor's requirement that no more than two hot spots per quadrant be assigned to intensive policing. This procedure assigned 756 hot spots to intensive policing and 1,163 to control.[42]

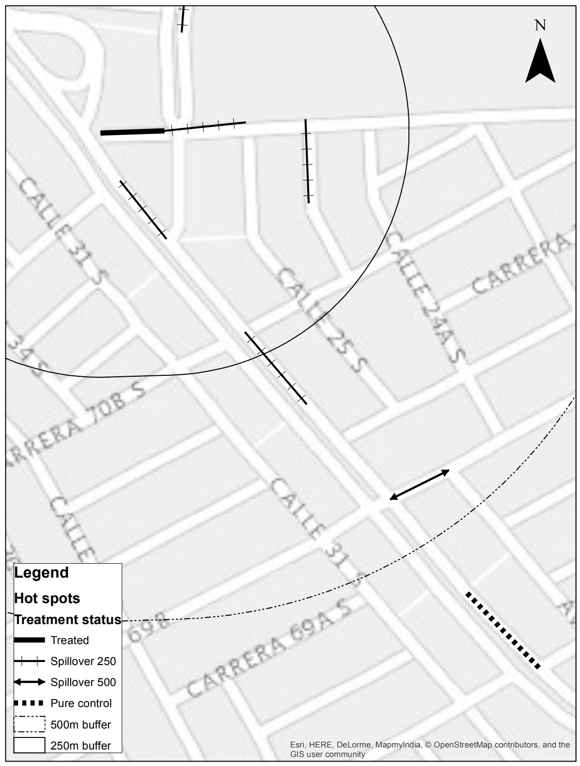
[39] For details on all pre-specified aspects of the design see https://www.socialscienceregistry.org/trials/1156. Previous literature on hot spots policing has focused mainly on catchment areas of about two blocks or 150 meters (Braga et al., 1999; Braga and Bond, 2008; Mazerolle et al., 2000; Taylor et al., 2011; Weisburd and Green, 1995). We felt 150 meters to be too conservative, however, and opted for 250 meters instead. We also specified a 500 meter option in case spillovers were unexpectedly large. Wider radii seemed implausible and would have eliminated the pure control category in a single city.

[40] If any share of a hot spot lies within the 250 meter radius of a treatment unit, we consider it to be in the proximate spillover region. Similarly, if any share of a hot spot lies within the 500 meter radius of a treatment unit we consider it to be in the distant spillover region.

[41] If criminals are relocating from treated to pure control segments, we should expect such non-spatial spillovers to overstate direct treatment effects and understate total spillovers; other biases are possible if reputation of increased government efficacy spreads non-spatially from treated to untreated areas. Uncertainty about how to tractably characterize the spillover process is a limitation of our approach, as it would be for virtually any evaluation of a place-based crime intervention.

[42] Within each station we took all quadrants with at least one hot spot and randomized quadrants to treatment with 0.6 probability. We then used complete randomization to assign hot spot segments to treatment within treatment quadrants. Specifically, in quadrants containing just one or two hot spots, we assigned all segments to treatment, while in quadrants with more than two hot spots we randomly assigned exactly two to treatment. This procedure is effectively a clustered random assignment, as all hot spots within some quadrants are jointly assigned to non-treatment while other segments in small quadrants are assigned jointly to treatment.

Figure 3: An example of assignment to the four treatment conditions



Notes: The figure depicts the classification scheme for streets across treatment and control conditions. We assign each segment in the experimental sample to either treatment or control condition. Then, we classify control segments into one of three experimental conditions according to their distance from the treated street: proximate spillover (<250 meters distant from a treated segment), distant spillover (250–500 meters), and "pure control" status (>500 meters). If any part of the segment falls inside the radius, it is considered in the corresponding control condition.

Table 3: Distribution of treatment and spillover assignments across the experimental sample

		Municipal services assignment to:					
		Treatment	<250m	250m-500m	>500m	*Ineligible*	All
Intensive policing assignment	Treatment	75	196	192	293	174	756
	<250m	74	281	185	165	162	705
	250m-500m	32	47	102	113	75	294
	>500m	20	22	16	106	49	164
	All	201	546	495	677	460	1,919

Notes: "Ineligible" segments are those that enumerators identified as having no garbage or broken lights. For simplicity, we ignore in this table whether ineligibles are proximal to hot spot policing or municipal services segments or not.

In March, after the Mayor's office decided to implement and evaluate municipal services, we selected streets for municipal services. We limited eligibility to the roughly 70% of segments that actually showed signs of physical disorder at baseline. We sent enumerators to take five photographs and rate hot spots for the presence of graffiti, garbage, boarded-up buildings, and run-down buildings.[43] Of the 1,534 segments they were able to safely visit, 30% had no need for maintenance and were excluded from eligibility, 65% reported either 1 or 2 maintenance issues, and 5% reported 3 to 5 issues. The 70% with at least one issue, and the 385 segments they could not visit safely, were eligible for municipal services assignment, for a total of 1,459 eligible hot spots. We blocked on police station and the previous intensive policing assignment, and assigned 201 hot spots (14% of eligible segments) to municipal services.[44]

Table 3 summarizes how the hot spot segments in our experimental sample are distributed across 20 treatment conditions and potential outcomes—4×5 conditions tied to the four conditions for each intervention (treatment, proximal, distant, and pure control) plus the ineligible category of streets that we deemed were in no need of municipal services.[45]

Tests of randomization balance Random assignment produced the expected degree of balance along covariates. Table 1 reports the weighted means for a selection of baseline covariates, by experimental assignment, for hot spots and non-hot spots, with additional covariates and tests in the Appendix.[46] For the most part, background attributes appear balanced across experimental conditions for the two pools of segments. There are some minor differences between treatment and control hot spots (for instance, treated hot spots are slightly less likely to be in industrial zones), but generally the imbalance is consistent with chance.

[43]One limitation of our approach is that this measure of baseline social disorder is measured after two months of the intensive policing treatment. We had little practical or theoretical reason to expect that increased police patrolling would affect physical disorder, and indeed there is no statistically significant difference between hot spot and non-hot spot streets. Enumerators were able to reach 1,534 of the segments, but because of personal safety concerns could not visit the remainder traveling on their own. Enumerators (who traveled individually or in pairs) determined which segments were safe and which were not on visual inspection, discussions with police at the CAI, or conversations with citizens. We instructed them to err on the side of caution, since these data on initial disorder were not crucial. According to our baseline administrative data, these unvisited segments had less overall crime and property crime, but more violent crime, in our baseline data. They were also farther away from social environments like churches and shopping centers.

[44]These 201 were the first "batch" to be treated. We also randomized a second batch of 214 hot spots for later treatment should the city decide to expand services. Two months into treatment of the first batch, however, our analysis of compliance records and visual inspection of hot spots suggested that continued municipal services were needed to maintain order in the first batch, and so the city did not give contractors the list of segments in the second batch. Thus the second batch remains in our control group.

[45]Technically there are 3 × 4 "ineligible" treatment conditions, since the streets that were diagnosed as having no need for municipal services could be <250 meters, 250–500 meters, or >500 meters from treatment streets, for both interventions.

[46]Appendix B.1 displays the results for all covariates and Appendix B.3 reports baseline balance for other pairwise comparisons, including treated segments to control streets excluding proximal spillover streets less than 250 meters away, plus spillover effects.

To see whether covariate imbalance lies within the expected range, we test the null hypothesis that the covariates do not jointly predict experimental assignment. We use multinomial logistic regression with randomization inference to model the four-category experimental assignments for hot spots or the three-category experimental assignments for non-hot spots.[47] The p-value is non-significant, as expected, for both the experimental and non-experimental (non-hot spot) samples. For hot spots, the log-likelihood statistic is 456.4, $p = 0.681$; for non-hot spots, the log-likelihood statistic is 19,097.0, $p = 0.531$.

5.3 Estimation

We estimate treatment and spillover effects within the *experimental sample* using the following weighted least squares regression:

$$Y_{sqp} = \beta_1 P_{sqp} + \beta_2 M_{sqp} + \beta_3 (P \times M)_{sqp} + \lambda_1 S^P_{sqp} + \lambda_2 S^M_{sqp} + \lambda_3 (S^P \times S^M)_{sqp} + \gamma_p + \Theta X_{sqp} + \epsilon_{sqp} \quad (1)$$

where Y is the outcome in segment s, quadrant q and police station p; P is an indicator for assignment to intensive policing; M is an indicator for assignment to municipal services; S^P and S^M are indicators for the relevant spillover region (either <250 meters or <500 meters from treatment, or a vector of both indicators). γ is a vector of police station fixed effects (our randomization strata); and X is a vector of pre-specified baseline control variables.[48] Weights are the inverse probability weights (IPWs) of assignment to each experimental condition(explained in more detail below).

To calculate spillovers in the *non-experimental sample (non-hot spots)* we estimate the following regression:

$$Y_{sqp} = \lambda^N_1 S^P_{sqp} + \lambda^N_2 S^M_{sqp} + \lambda^N_3 (S^P \times S^M)_{sqp} + \gamma_p + \Theta X_{sqp} + \epsilon_{sqp} \quad (2)$$

using IPW for assignment to the conditions S^P and S^M based on proximity to all hot spots in the experimental sample.

Thus, β_1 and β_2 estimate the marginal effects of each treatment, and β_3 estimates the interaction effect of receiving both. A negative sign on β_3 would imply increasing returns, and a positive sign decreasing returns. The cumulative effect of receiving both interventions is the sum of the three coefficients, $\beta_1 + \beta_2 + \beta_3$. Likewise, the λ and λ^N estimate spillover effects of each treatment in the experimental and non-experimental samples. If we wish to see the simple marginal effects of each treatment, we can estimate each equation with the constraints that $\beta_3 = 0$ and $\lambda_3 = 0$. These constraints are important especially for those analyses (such

[47] Ordinarily, we would take the p-value of the log-likelihood statistic from a chi-square distribution. In this case, the log-likelihood statistic's distribution under the null hypothesis is unknown due to the complications posed by clustered random assignment of treatments and geographic clustering in the assignment of spillovers. To obtain exact p-values, we instead use randomization inference. Using simulated random assignments, we obtain a reference distribution of log-likelihood statistics under the null hypothesis; we then calculate the p-value by locating the actual log-likelihood value within this reference distribution.

[48] We selected these covariates by measuring their ability to predict baseline crime levels. To measure the prognostic ability of this baseline data, we regressed the 2015 crime index on an indicator for treatment and each of the combinations of covariates for 100 different treatment assignments. These included pre-treatment administrative data as well as pre-treatment crime and patrolling measures. We weighted the observations by the inverse of the probability of being in its observed experimental condition and restricted the sample to only hot spots with a non-zero probability of being assigned to both treatment and control. We examined where the empirical standard error from the all covariates version landed in the distribution of empirical standard errors from all other sets of covariates. Our analysis revealed that including all covariates produced an empirical standard error in the bottom five% of all empirical standard errors, so we include all covariates in our regressions since it may be that the optimal covariates slightly outperform all covariates because of chance. We pre-specified this covariate selection procedure. The control vector X also includes an indicator for segments that were ineligible for municipal services treatment by virtue of their baseline disorder.

as the first stage estimates of treatment compliance) where theoretically we expect no interaction.[49]

Inverse probability weighting Spillovers introduce another source of spuriousness that can be corrected with IPWs. Hot spots close to other hot spots, such as those in the city center or other dense areas, will be assigned to the spillover condition in most randomizations. These streets may have unobservable characteristics that are associated with high levels of crime. This can be a source of spuriousness in estimating spillover effects, and could mechanically lead us to conclude that spillovers increase crime. Controlling for baseline characteristics and crime histories reduce but do not eliminate the potential bias. With IPWs, outcomes for the segments assigned to any given condition are weighted by the inverse of the probability of assignment to that condition.[50] In general, they deflate the importance of streets that have a high probability of assignment to a given experimental condition, and ensure that all segments have the same probability (after weighting) of getting the spillover treatment.

Procedure for determining the spillover condition To determine the relevant spillover radii for conditions S^P and S^M, we pre-specified a procedure: if there is no evidence of statistically significant spillovers into the 250–500 meter region (using a $p < .1$ threshold), then the conditions S^P and S^M will indicate segments in the <250 meter spillover region only, otherwise they will indicate segments <500 meters of treated hot spots. If there are no statistically significant spillovers in the 250 meter radius nor the 250-500 meter radius, then our primary estimates would ignore the classification of control streets into various spillover conditions and estimate the β coefficients alone.

In retrospect, our pre-specified rule for determining the spillover range was probably too permissive in two respects. First, it was based on spillovers in the experimental sample rather than the much larger non-experimental sample. Second, this rule could lead us to ignore quantitatively large but imprecise spillovers. In principle it could lead us to ignore spillovers large enough to outweigh any direct benefits of crime reductions in treatment hot spots. Thus we will show results accounting for spillovers considering the non-experimental sample of streets and estimate aggregate effects based on the number of streets exposed to spillovers.

Estimating standard errors In the simple no spillovers case, we can estimate accurate standard errors clustering by randomization strata.[51] As soon as we include the S^P and S^M conditions, however, we introduce difficult-to-model patterns of clustering. Conventional standard errors will be too small. The next

[49]This estimation strategy represents a slight departure from the pre-analysis plan. The plan indicated that we would first and foremost focus on pairwise comparisons of each intervention separately, dropping from the regression any segments with a zero probability of assignment to any of the conditions. That approach generates similar results but, in retrospect, is problematic. Most importantly, a pairwise comparison of streets that did and did not receive intensive policing (ignoring municipal services treatment) would be biased since assignment to municipal services is slightly imbalanced across intensive policing experimental conditions (see Table 1). Hence we must estimate the effects of both interventions jointly. In addition, our original approach required us to drop an increasing number of segments from the regression, especially when estimating the interaction, rather than using the full sample. Equations **??** and 1 maintain the spirit of the original estimation approach but correct for these problems.

[50]Each segment's probability of exposure to proximal or distant spillovers can be estimated with high precision by simulating the randomization procedure a large number of times. Such inverse probability weights have a long history in survey sampling and have become common in the analysis of randomized trials with varying probabilities of assignment (Horvitz and Thompson, 1952; Gerber and Green, 2012). Appendix C.3 reports estimates of treatment effects without these weights.

[51]Technically, we use the following rules that reflect our randomization approach: (i) for treated segments in quadrants with more than two or exactly one hot spot, each segment is a cluster; (ii) for other treated segments (those in quadrants with two hot spots, both assigned to treatment), the quadrant is a cluster; (iii) for control hot spots in treatment quadrants, each segment is a cluster; and (iv) for control hot spots in control quadrants, the quadrant is a cluster. Note that our randomization approach first assigned quadrants to treatment or control. In treatment quadrants with one or two hot spots, we assigned both streets to treatment, while in treatment quadrants with more than two hot spots, we randomly assigned two to treatment. In control quadrants, we assigned all hot spots to control.

section discusses these simulations and the randomization inference approach.

Continuous spillovers Overall, the above approach is similar to the approach that previous studies have used to estimate spillovers into a nearby catchment area. Our advantages include: we can estimate spillovers flexibly over various radii; we can account for overlapping catchment areas of both the hot spots policing intervention and the municipal services intervention; and we can estimate exact p-values. Alternatively, we can estimate a continuous, monotonic spatial decay function with the following OLS regression:

$$Y_{sqp} = \breve{\beta}_1 P_{sqp} + \breve{\beta}_2 M_{sqp} + \breve{\lambda}_1 \sum_{t \in T_P} f(d_{sqp,t}) + \breve{\lambda}_2 \sum_{t \in T_M} f(d_{sqp,t}) + \breve{\gamma}_p + \breve{\Theta} X_{sqp} + \epsilon_{sqp} \tag{3}$$

where $f(d_{sqp,t})$ is a spatial decay function with a standardized distribution. It is a weighted sum of distances to all treated hot spots, where t enumerates treated hot spots and T is the set of all treated hot spots. Treated segments receive no spillover from themselves but can receive spillovers from other treated segments. Applied to the non-experimental sample, the regression would omit the direct treatment effects. Our default functional form is exponential, $f(d_{sqp,t}) = 1/(e^{d_{sqp,t}})$, but we examine alternatives. We can no longer employ IPWs to weight street segments because the exposure measures are continuous variables and do not have a finite number of outcomes. Instead, we include in the control vector expected spillover intensities (the averages across 1,000 simulations) and the probabilities of being treated by each intervention. As before, we calculate standard errors using randomization inference.

5.4 Why randomization inference?

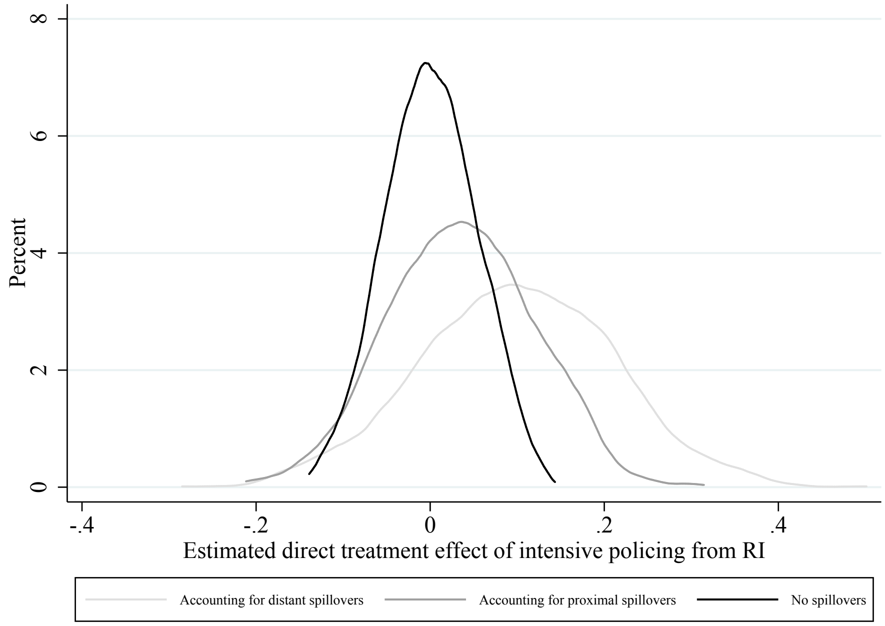
Randomization inference (RI) gives precise p-values based on the empirical distribution of all treatment effects that could arise under our specific design and data. RI reassigns treatment randomly thousands of times, each time estimating the treatment effect that could have arisen by chance from that specific comparison. Figure 4 displays these empirical distributions for three cases of equation (1): the simple treatment-control comparison with no spillovers (i.e. $S^P_s = S^M_s = 0$ for all s); the case where S^P and S^M indicate proximal spillovers within 250 meters; and the case where S^P and S^M indicate the larger spillover area within 500 meters.

Most importantly, note that the distribution widens when accounting for spillovers. The no spillovers case has the narrowest distribution, and the p-values associated with each treatment effect are nearly identical to the p-values obtained from the conventional WLS standard errors clustered by randomization strata. The distribution widens as we account for further and further spillovers. In short, we are more likely to get large treatment effects by chance, in both directions. Thus the standard errors estimated by the WLS regressions 1 and 2 will tend to be too small.

This widening of the sampling distributions follows from two facts. One is that we are losing data as we pare off rings of spillovers. The second is that the control region becomes more separate from the region where the hot spots are located. Hot spots that are close to other hot spots are assigned to the spillover condition in most randomizations, creating patterns of clustering. These clusters are difficult to model because they have to do with general distance from other hot spots rather than an easily defined characteristic such as a geographic area or stratum.

The simulations in Figure 4 also point to another identification problem. Estimating spillovers using equations 1 and 2 can lead to a small level of bias in estimated coefficients, even when using IPWs. Clustered assignment introduces bias when there are clusters of unequal size, and when cluster size is correlated with

Figure 4: The empirical distribution of estimated treatment effects on insecurity under different spillover scenarios



Notes: The figure displays the empirical distribution of treatment effects for intensive policing depending on the number of spillover units defined. The dependent variable is the standardized insecurity index. To generate the distributions, we simulate the randomization procedure 1,000 times and estimate treatment effects for each randomization using end-line data under the sharp null of no treatment effect for any unit. The figures show distributions for three cases of equation (1): the simple treatment-control comparison with no spillovers (i.e. $S_s^P = S_s^M = 0$ for all s); the case where S^P and S^M indicate proximal spillovers within 250 meters; and the case where S^P and S^M indicate the larger spillover area within 500 meters. Under the sharp null, the direct treatment effect of intensive policing should be zero for every unit because the outcome variable is the same regardless of the unit's treatment assignment. However, accounting for proximal or distance spillovers generates distributions that are not centered at zero because of clustering from the randomization procedure.

potential outcomes. When we ignore spillovers, we stipulate that there is no such clustering, which is why that distribution is centered at zero. When we allow for spillovers, we confront the fact that our exposure to spillovers is clustered. The bias disappears as the number of clusters increases (and indeed it is negligible when we estimate non-experimental spillovers). Fortunately, the bias on direct effects is not large. Unfortunately, the actual direct effects we estimate will often be subtle, and so the bias is fairly large in comparison to some of the direct average treatment effects.

What RI allows us to do is to assign a p-value for a given treatment effect by observing where that treatment effect falls in the distribution of all possible effects in 10,000 randomizations. We use these RI p-values in place of the conventional standard errors and p-values whenever we estimate treatment effects in the presence of spillovers. RI also estimates the bias. All of our tables report bias-corrected treatment effects. Appendix B.7 reports the specific biases estimated. Finally, Appendix C.3 reports estimates of treatment effects without weights and randomization inference, and we discuss these results below.

6 Results

6.1 Program implementation and compliance

Broadly speaking, the police patrols and municipal services complied with the Mayor's instructions and treatment assignment. Police did so for the full eight months, while municipal services agencies likely complied for a shorter period. Table 4 reports the effects of assignment to intensive policing or municipal services on a variety of first-stage outcomes. We estimate equation 1 ignoring interactions between the two treatments since we have no reason to expect one treatment to affect compliance with another. For simplicity we compare treatment segments to all control segments, ignoring spillovers, but draw similar conclusions accounting for spillovers (see Appendix C.1).

Intensive policing Calculating the time spent on street segments is difficult because of periodically malfunctioning units or outages, plus variation in data quality.[52] We estimate control streets received 86 minutes of patrolling time per day, on average. Treated streets received an extra 65 minutes, a 75% increase.[53] Streets outside the experimental sample received an average of 33 minutes of patrolling time per day.[54]

Without pre-treatment data on patrol times it is impossible to say whether the increase in patrol time on treatment hot spots came at the expense of control hot spots. What we can say is that the 65 minute rise on two segments means roughly a minute less time on each of the 130 other segments in the quadrant.

We do not see any effect of increased policing on arrests or police actions such as drug seizures, suggesting any effect of the policing may be through deterrence rather than incapacitation (Chalfin and McCrary, 2017).

Citizens seemed to have noticed an increase in patrols, albeit imperfectly. The survey asked whether citizens thought patrolling on the segment increased, decreased, or stayed the same in the past 6 months. On control segments 13% reported an increase, compared to 21% on treatment segments.

Municipal services Table 5 summarizes compliance. After we assigned the 201 streets to municipal services, city agencies visited each street for a formal diagnostic. They identified 123 streets needing and eligible for clean-up services, and 47 needing lighting improvements. They conducted the diagnostics in March and performed the services in June through August. Tree pruning and graffiti cleaning were supposed to be one-time treatments; rubbish collection was expected to be semi-regular. Based on the city's administrative data, 74 of the 123 streets (60.2%) were treated with either tree pruning or rubbish collection or both, and in

[52] We estimated patrolling time using the time stamp of the GPS pings sent by every device. In the easiest cases, several sequential pings were received from the area of 40 meters surrounding a segment. In this case, we took the first ping as the entry time and the last as the exit time, and computed the patrolling time for an entry. Then, we aggregated entries to measure daily patrolling times. However, because of malfunctioning units, there were several cases in which irregular and largely separated pings were sent by a device. To account for these situations, we top-coded each entry up to the duration of the shift (starting with the entry time). We also drop days with missing data, as it was more likely that the device was not working than the street was not patrolled at all during the day. We discussed these adjustments with the police to ensure we were making a correct approximation of daily patrolling times.

[53] If we account for potential spillovers, we do not see significant spillovers: streets within 250 meters of treated hot spots received about 8 more minutes of patrol time compared to more distant segments ($p = 0.34$). See Appendix C.1. One exception to the increase in patrolling time was the Bronx, three streets wholly controlled by organized crime, which had no police presence prior to the study, and over the course of the study was invaded by a especial branch of police, cleared of all population, and scheduled for demolition. Two were treatment streets and one was a control street. They are de facto dropped from the sample because there was no crime data or ability to survey after the operation. Hence our sample size in all tables is 1,916 not 1,919.

[54] One concern on measuring police compliance was that treatment quadrants may be better equipped than other quadrants and therefore would have improved measures of patrolling time. To assess this situation, we compare the distribution of quadrant $\times$ days with no GPS pings between treatment and control quadrants. We find no statistically significant differences in the number of quadrant $\times$ days with no GPS pings between both groups. See Appendix B.8.

Table 4: First-stage effects of treatment assignment on measures of compliance and effectiveness

Dependent variable	Control mean	ITT of assignment to: Intensive policing	Municipal services
	(1)	(2)	(3)
A. Intensive policing measures:			
Proportion of respondents who say police presence increased in last 6 mo.	0.129	0.076	0.017
		[.011]***	[.013]
Daily average patrolling time, excluding quadrant-days without data	92.001	76.571	-3.333
		[4.424]***	[4.371]
# of arrests	0.333	-0.053	0.026
		[.082]	[.102]
# of drug seizure cases	0.041	-0.002	0.029
		[.020]	[.024]
# of gun seizure cases	0.009	0.006	0.007
		[.008]	[.013]
# of recovered car cases	0.003	0.000	-0.003
		[.001]	[.001]*
# of recovered motorbike cases	0.006	-0.028	0.032
		[.019]	[.027]
B. Municipal services implementation measures			
Proportion of respondents who say municipal presence increased in last 6 mo.	0.144	0.006	0.016
		[.010]	[.012]
City determined segment is eligible for lights intervention	0.349	-0.007	-0.139
		[.048]	[.048]***
Received lights intervention	0.000	-0.010	0.199
		[.020]	[.026]***
City determined segment is eligible for garbage intervention	0.000	0.011	0.627
		[.025]	[.032]***
Received garbage intervention	0.000	0.015	0.382
		[.026]	[.033]***
June 2016 enumerator assessment of street conditions:			
Graffiti on segment	0.749	-0.018	0.078
		[.050]	[.043]*
Garbage on segment	0.251	0.071	0.015
		[.061]	[.049]
Visibly broken street light on block	0.000	0.012	0.008
		[.012]	[.008]
December 2016 enumerator assessment of street conditions:			
Graffiti on segment	0.624	0.019	0.059
		[.053]	[.047]
Garbage on segment	0.245	0.021	0.002
		[.051]	[.043]
Visibly broken street light on block	0.029	0.022	-0.015
		[.016]	[.017]

Notes: This table reports intent to treat (ITT) estimates of the effects of the two interventions, via a WLS regression of each outcome on treatment indicators, police station (block) fixed effects, and baseline covariates (see equation 1, where we have constrained the coefficient on the interaction term to be zero and ignored spillovers). The regression ignores spillover effects. Standard errors are clustered using the following rules: (i) for all treated segments except with cluster size 2, each segment is a cluster; (ii) for all other untreated segments, each segment gets its own cluster identifier; (iii) for entirely untreated quadrants, they form a cluster; and (iv) for quadrants with exactly 2 units assigned to treatment, those units form a cluster. The proportion of people reporting increased state presence comes from our citizen survey, the enumerator assessments were collected by the research team, and the remainder of the outcomes come from police administrative data. * significant at the 10 percent, ** significant at the 5 percent, *** significant at the 1 percent.

Table 5: Municipal services eligibility and compliance

		City's lighting assessment			% of eligible streets
		Lights eligible	Lights ineligible	All	receiving lighting service
City's	Eligible for garbage	21	102	123	74 (89.1%)
cleanliness	Ineligible for garbage	26	52	78	
assessment	All	47	154	201	
Eligible streets receiving clean-up		41 (60.2%)			

Notes: The table summarizes compliance on the municipal services intervention for 201 streets assigned to treatment as reported by the corresponding agencies within the Mayor's office.

41 of the 47 streets (87%) they repaired broken lights and replaced older and less bright lights with modern brighter ones. No graffiti was cleaned-up.

Despite this maintenance work, the impacts were not necessarily visible. Based on our survey data in Table 4, 14.4% of respondents on control segments noticed an improvement in service delivery in the past six months, and this was only 1.9 percentage points greater in treatment streets (not statistically significant). We also visited segments in the daytime in June and December 2016 to photograph and have enumerators record their impressions of the streets. The before and after photos generally display streets with relatively minor levels of litter and open waste, and relatively few plants in need of major pruning. Before-after differences are nearly imperceptible. It is possible that lights repairs were more evident, but it was unsafe for our enumerators to visit hot spots at night. Enumerator ratings of street cleanliness are reported in Table 4, and we see no effect of treatment. One possibility is that the extensive margin is the wrong margin to evaluate, and another is that the disorder in after photos could have accumulated over days or weeks.

6.2 Program impacts on officially reported crime

We begin by analyzing impacts using administrative crime data from all streets in the city. Table 6 reports results from estimating the direct treatment (β), experimental spillover (λ), and non-experimental spillover (λ^t) coefficients from equations (1) and (2), with and without the interaction terms between intensive policing and municipal services.[55] Table 6 estimates spillovers within 250 meters only. This follows our pre-specified rule given that, as we will illustrate below, we do not see statistically significant evidence of spillovers in the 250–500 meter region. Appendix C.8 reports alternative spillover regions and approaches.

Table 6 also calculates the total number of deterred crimes, as the product of the estimated coefficients and the number of treatment and spillover segments in the city. There are many more spillover streets than treated streets—51,390 non-hot spots and 705 control hot spots for the policing intervention and 20,740 non-hot spots and 546 control hot spots for the municipal services intervention. We drop the 57,695 streets with zero probability of assignment to the spillover condition from this analysis (i.e. more than 250 meters from a hot spot in the experimental sample). Thus even small estimated spillovers can have a large effect on the total crime estimates. Since all our coefficients are fairly uncertain, we have to take these aggregate impacts with some caution.

Our best guess for the overall impact on crime is that the interventions directly deters a relatively modest amount of crime, and that some or all of this crime is displaced to neighboring streets. However, as we show below, crime displacement is concentrated in property crime. The data suggests that violent crimes may not

[55] We pre-specified a one-tailed test for these regressions, since we had strong priors about the direction of the effect, but significance levels in the table reflect a two-tailed test to be conservative and consistent with the other analyses.

be displaced so easily.

Direct treatment effects Starting with columns 1–4 (no interaction between treatments), both intensive policing and municipal services reduce officially reported crimes on average, although these coefficients are not statistically significant. Control segments report an average of 0.743 crimes over the intervention period. Thus the coefficient on intensive policing of -0.094 represents a 12.6% improvement. The municipal services coefficient is about two-thirds as large. In total, these estimates suggest that the reallocation of police and municipal services deterred a total of 86 crimes in targeted streets over the intervention period. Of course this estimate is extremely imprecise.

Turning to columns 5–8, we see larger and most statistically significant impacts of state presence in the segments that received both interventions. The coefficients on intensive policing and municipal services are weakly positive. We see no evidence that either intervention on its own reduced crime. The coefficient on the interaction is -0.437, however, with a RI p-value of 0.043. The sum of the three coefficients is -0.339 with a p-value of 0.109. This sum corresponds to a 45.6% decrease in reported crimes on the 75 streets that received both interventions. The fact that the coefficient on the interaction is large, negative, and statistically significant implies that there may be increasing returns to security investments, at least over this range of variation. Of course, given that the sum of effects is weakly statistically significant, we cannot conclude with a high degree of confidence that both interventions together reduced crime on these 75 streets. Moreover, the aggregate direct effect of the program looks even smaller when we account for the interaction. According to these estimates, our best guess is that only 8 crimes were deterred directly by both interventions—about 1 per month of the policing and municipal services.

Spillover effects Meanwhile, the spillover coefficients suggest that any crime deterred is more than made up for by a rise in crime in streets within 250 meters from targeted hot spots. For intensive policing, all four spillover coefficients are positive, pointing to a displacement of crime to nearby streets. The spillover effects in the experimental sample are imprecise, but given the large number of nearby non-hot spots, the spillovers in the non-experimental sample are significant at almost the 10% level when we do not allow for the interaction between treatments. There are a sufficiently large number of non-hot spot segments that these small coefficients add up to high levels of crime—more than 800 crimes in aggregate when we do not allow for the interaction and more than 600 when we do. In contrast, we see no evidence that municipal services push crime around the corner. The coefficients on spillovers in the non-experimental sample are actually negative, though they are imprecise. In aggregate, however, this estimate adds up to between 50 and 100 crimes deterred in nearby streets, depending on the specification.

Aggregate effects citywide We use these estimates to guess the aggregate effect on crime. We can conclude with some confidence that reallocating police and municipal services did not reduce crime levels in the city. Indeed, the calculations point to the opposite. The estimates suggest crimes increased by about 800 in both specifications—with and without the interaction. This must be taken with caution, however, as we generate confidence intervals for these totals using randomization inference and neither aggregate effect is statistically significant at the 5% or 10% levels.[56] We are skeptical that crime increased in aggregate by so

[56] First we create a fake schedule of potential outcomes for each observation by adding or subtracting RI-adjusted treatment or spillover effects. This process gives us a potential outcome for each unit depending on its treatment assignment. Second, we simulate a randomization and take the potential outcome associated with the treatment assignment of the new randomization. Third, we estimate treatment and spillover effects using this new outcome and apply the RI bias adjustment from our main set of results. Fourth, we multiply these bias-adjusted treatment effects by the number of segments in each group, and sum across

Table 6: Estimated aggregate impacts of the interventions, accounting for proximal spillovers in the experimental and non-experimental samples

	Dependent variable: # of crimes reported to police on segment (administrative data)							
	No interaction between treatments				Interaction between treatments			
Impacts of treatment	Coeff. (1)	RI p-value (2)	# segments (3)	Estimated total impact = (1) × (3) (4)	Coeff. (5)	RI p-value (6)	# segments (7)	Estimated total impact = (5) × (7) (8)
A. Direct treatment effect								
Intensive policing	-0.094	*0.512*	756	-70.7	0.009	*0.817*	756	7.2
Municipal services	-0.076	*0.782*	201	-15.2	0.089	*0.367*	201	17.9
Both					-0.437	*0.043*	75	-32.8
Subtotal				-85.9				-7.8
B. Spillover, experimental sample								
Intensive policing	0.061	*0.595*	705	42.7	0.143	*0.315*	705	100.6
Municipal services	0.176	*0.056*	546	96.3	0.255	*0.025*	546	139.2
Both					-0.272	*0.196*	281	-76.5
Subtotal				138.9				163.3
C. Spillover, non-experimental sample								
Intensive policing	0.016	*0.113*	51390	840.7	0.013	*0.219*	51390	654.1
Municipal services	-0.003	*0.416*	20740	-55.5	-0.006	*0.500*	20740	-120.8
Both					0.006	*0.968*	15491	95.7
Subtotal				785.2				629.0
Net increase (decrease) in crime				838.2				784.5
			95% CI	(–813, 2131)			*95% CI*	(–1063, 2268)
			90% CI	(-492, 1919)			*90% CI*	(-735, 2033)

Notes: This table presents the aggregate effect calculation for both interventions assuming proximal spillovers. Columns 1–4 refer to the non-interacted results (equation 1 under the constraint that $\beta_3 = 0$ and $\lambda_3 = 0$) while columns 5–8 refer to the interacted results (equation 1 with no constraints). Columns 1 and 5 display the bias-adjusted treatment effect while columns 2 and 6 display RI p-values. Columns 3 and 7 display the number of units in each group. Columns 4 and 8 display the product of the bias-adjusted treatment effect and the number of units in each group. The confidence interval on the bottom of the table is constructed using randomization inference. First we create a fake schedule of potential outcomes for each observation by adding or subtracting RI-adjusted treatment or spillover effects. This process gives us a potential outcome for each unit depending on its treatment assignment. Second, we simulate a randomization and take the potential outcome associated with the treatment assignment of the new randomization. Third we estimate treatment and spillover effects using this new outcome and apply the RI bias adjustment from our main set of results. Fourth, we multiply these bias-adjusted treatment effects by the number of segments in each group, and sum across both the experimental and non-experimental samples to get the aggregate effect. We repeat steps two through four 1,000 times to get the distribution of the test statistic, which is roughly centered on the actual number of deterred crimes. The 2.5 and 97.5 percentiles of this distribution give us the 95% confidence interval. $p < .1$ in bold.

Table 7: Aggregate impacts on crimes by type (mean and confidence intervals)

	without interaction			*with interaction*		
	Effect	95% CI	90% CI	Effect	95% CI	90% CI
	(1)	(2)	(3)	(4)	(5)	(6)
All crime	838.2	(–813, 2131)	(-492, 1919)	784.5	(–1063, 2268)	(-735, 2033)
Property crime	1014.4	(-195, 2075)	(-44, 1903)	1205.1	(-340, 2385)	(23, 2239)
Violent crime	-135.0	(-853, 389)	(-747, 281)	-374.1	(-1134, 213)	(-1011, 75)
Homicides and sexual assaults only	-65.3	(-178, 55)	(-162, 41)	-97.1	(-236, 32)	(-210, 16)
Property – violent crime	1149.3			1579.1		
p-value	0.068			0.018		

Notes: This table presents the aggregate effect calculation for various crime subgroups assuming proximal spillovers. Calculations are based on the aggregate effect and confidence interval described in Table 6

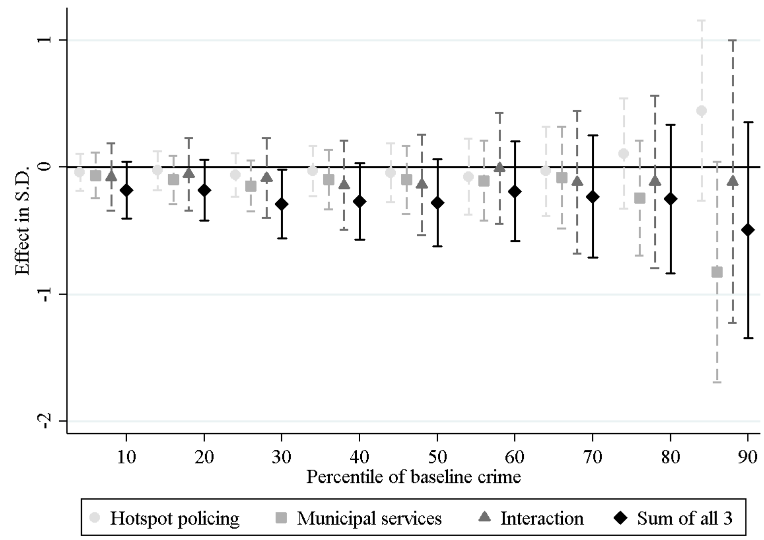
much, and in the conclusions section discuss what could explain these patterns. But these estimates suggest that we can rule out the possibility that crime decreased in the city by even a modest amount. However, as we explain below, this result is driven mainly by property crimes.

Heterogeneity by type of crime These crime totals conceal important heterogeneity by crime type. Police generally distinguish between violent crimes (murder, rape, assaults, threats) and property crimes (such as burglary or car theft). Violent crimes tend to be of greatest concern to authorities, especially homicides and sexual assaults. Table 7 disaggregates the impacts on total crime into violent and property crimes. Based on these estimates, the best guess is that aggregate violent crimes fell by 135 to 374 crimes in total, depending on whether we use the interaction or not, although neither estimate is statistically significant. Property crimes rose by 1,014 to 1,205 in aggregate, however, and these estimates are statistically significant at the 10% level when we include the interaction. If we focus on the most two socially costly crimes for which we have data (homicides and sexual assaults), we see incidents fall by 65 to 97. What is more, the difference in property and violent crimes is statistically significant.

Heterogeneity by initial level of crime We pre-specified one major form of heterogeneity analysis, by baseline levels of crime. Broadly, we observe what we predicted: that improvements in insecurity are greater in the higher-crime streets. Figure 5 reports the results of estimating equation 1 nine times, each time interacting each treatment indicator with an indicator for whether a segment is below the nth percentile of baseline crime levels among our experimental sample of hotspots, for $n = 10, 20, \ldots, 90$. The coefficients on the treatment indicators indicate the effect on the higher crime segments above that percentile. The figure graphs these coefficients. The treatment effect is fairly constant up until the point we reach the street segments in the 90^{th} percentile and above, when the impact of receiving both interventions climbs to 0.5 standard deviations. The effect is imprecise, as the sample size drops dramatically. For these streets, the marginal effect of receiving the municipal services treatment is on the margin of statistical significance, however the small sample. These results are consistent with increasing returns to treating the less secure hot spots. The cumulative effect of both interventions appears within 8 to 12 weeks of the intervention, and grows over time (see Appendix B.10.3).

both the experimental and non-experimental samples to get the aggregate effect. We repeat steps two through four 1,000 times to get the distribution of the test statistic, which is roughly centered on the actual number of deterred crimes. The 2.5 and 97.5 percentiles of this distribution give us the 95% confidence interval, while the 5 and 95 percentiles give us the 90% confidence interval.

Figure 5: Heterogeneity of security impacts by pre-treatment administrative crime levels



Notes: The figure reports the ITT effect of the two interventions and the interaction term, plus the sum of these three coefficients, where the sample includes all segments to the right of each percentile (as opposed to looking only within deciles of baseline crime).

6.3 Program impacts on insecurity, including survey-based outcomes

Table 8 reports program impacts on the two main security measures described above: (i) the perceived risk index, based on survey responses; and (ii) the index of crime, which averages survey-based reports of incidence and the officially reported crime measure examined in the previous section. Both measures are standardized, and so treatment effects can be interpreted as average standard deviation changes in the outcome index. The table also reports treatment effects on the two components of the crime index, for transparency. While our focus is on the two pre-specified indexes, perceived risk and crime, to reduce comparisons further we also report results for an equally-weighted average of the perceived risk and crime incidence indexes.[57]

The table estimates equation (1), and reports both direct treatment effects and spillover effects on hot spots within 250 meters. Appendix C.10 reports robustness to alternative spillover regions.

Broadly speaking, we draw similar conclusions from the survey data as the administrative data on reported crimes. We see the largest and most statistically robust impacts of state presence in the segments that received both interventions. Those 75 segments reported a 0.327 standard deviation decrease in overall insecurity, significant at the 10% level (Column 5, which reports the sum of the the marginal effects, $\beta_1 + \beta_2 + \beta_3$). The coefficients on perceived risk and crime indexes are similar, though only the perceived risk index is statistically significant alone.

Alone, the interventions are associated with improvements in security, but none of the estimates are statistically significant on their own. Nonetheless the coefficients uniformly point in the direction of better security: intensive policing alone reduces perceived risk by 0.12 standard deviations, crime by 0.06, and overall insecurity by 0.11; while municipal services alone reduces perceived risk by 0.09 standard deviations,

[57]We did not pre-specify this aggregate “insecurity index” but it is a useful summary measure to focus on. Appendix B.10.1 reports treatment effects on the components of these summary indexes.

Table 8: Program impacts on security in the experimental sample, accounting for spillovers within 250 meters, with p-values from randomization inference (N=1,916)

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both	Sum of (2), (3), and (4)	Impact of proximal spillovers: Any intensive policing	Any municipal services	Both	Sum of (6), (7), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Insecurity index, z-score (+ more insecure)	-0.003	-0.106	-0.100	-0.121	-0.327	0.045	0.150	-0.217	-0.022
		0.391	*0.536*	*0.447*	***0.095***	*0.322*	***0.020***	***0.039***	*0.577*
Perceived risk index, z-score (+ riskier)	0.049	-0.122	-0.086	-0.084	-0.292	0.002	0.083	-0.160	-0.075
		0.259	*0.494*	*0.644*	***0.094***	*0.511*	*0.129*	***0.085***	*0.808*
Crime index, z-score (+ more crime)	-0.054	-0.054	-0.080	-0.118	-0.252	0.073	0.166	-0.200	0.039
		0.701	*0.659*	*0.412*	*0.196*	*0.231*	***0.010***	***0.059***	*0.361*
Perceived & actual incidence of crime, z-score	0.059	-0.081	-0.158	0.066	-0.173	0.027	0.099	-0.137	-0.011
		0.514	*0.153*	*0.423*	*0.507*	*0.418*	***0.092***	*0.171*	*0.578*
# crimes reported to police on street segment	0.743	0.009	0.089	-0.437	-0.339	0.143	0.255	-0.272	0.125
		0.817	*0.367*	***0.043***	*0.109*	*0.315*	***0.025***	*0.196*	*0.289*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports intent to treat (ITT) estimates of equation 1, estimating the direct effects of the two interventions (Columns 2 to 4) and the spillover effects (Columns 6 to 8) via a WLS regression of each outcome on treatment indicators, spillover indicators, police station (block) fixed effects, and baseline covariates. Columns 5 and 9 report the sum of the three preceding coefficients. The measures of perceived risk, perceived incidence of crime, and proportion reporting crime come from our citizen survey, and the # of crimes reported to the police come from police administrative data.

Table 9: Security impacts on non-experimental street segments <250m from treatment hot spots (N=399)

Dependent variable	Control mean	Impact of proximal spillovers: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)
	(1)	(2)	(3)	(4)	(5)
Insecurity index, z-score (+ more insecure)	-0.290	0.112	-0.002	-0.255	-0.145
		0.415	*0.966*	*0.269*	*0.435*
Perceived risk, z-score (+ riskier)	-0.099	0.018	-0.131	-0.136	-0.249
		0.925	*0.470*	*0.616*	*0.156*
Crime incidence, z-score (+ more crime)	-0.383	0.169	0.129	-0.289	0.009
		0.134	*0.372*	*0.154*	*0.822*
Perceived incidence of crime, z-score	-0.152	0.185	0.140	-0.270	0.055
		0.219	*0.478*	*0.304*	*0.741*
# crimes reported to police on street segment	0.271	0.096	0.076	-0.253	-0.081
		0.336	*0.407*	*0.167*	*0.826*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports spillover effects in the non-experimental sample from equation 2, a WLS regression of each outcome on spillover indicators, police station (block) fixed effects, and baseline covariates 1. In panel (b), Column 5 reports the sum of the three spillover coefficients.

crime by 0.08, and overall insecurity by 0.10. Also, the coefficient on the interaction term (Column 4) is not statistically significant for any of the survey measures or the overall measure, implying we cannot say with confidence that the hot spots treated with both interventions are different than the hot spots treated with just one. However, the interaction is significant for officially reported crimes. We take this result as being suggestive of increasing returns to state presence.

Spillovers There is also evidence of crime displacing to control hot spots. Columns 6 to 9 of Table 8 report these estimates. Intensive policing alone and municipal services alone are associated with increases in crimes on nearby hot spots of 0 to 0.25 standard deviations. Only the municipal services impacts are statistically significant, with a 0.15 standard deviation increase in insecurity. The interaction terms are generally negative (Column 8) and generally statistically significant, such that there is generally no evidence of spillovers onto hot spots near to hot spots that received both intensive policing and municipal services. When we ignore the interaction between treatments, we still find statistically significant evidence of spillovers resulting from the municipal services intervention (see Appendix C.4).

We also have survey data on 399 non-experimental street segments, and Table 9 estimates these non-experimental spillovers within 250 meters using equation (2). This sample is generally too small to estimate non-experimental spillovers precisely, but the patterns are generally consistent with what we see in the large-sample dataset on reported crimes, in Table 6. The coefficients on intensive policing are positive. The coefficients on municipal services vary, but the sign on the index of overall insecurity is negative (and extremely close to zero). Unlike the effects on reported crime in the large sample, the coefficients on the interaction terms are generally negative. We report results using different spillover regions in Appendix C.10.

Disentangling municipal services Some of our qualitative work and compliance data hinted that the lighting intervention may have been more compliant, effective, and persistent than the street clean-up aspects

Table 10: Municipal services impacts by subgroup

Independent variable	Dependent variable: Index of insecurity (z-score)			
		Block 1 versus Block 2		
	Full sample	All	Lights eligible only	Lights ineligible only
	(1)	(2)	(3)	(4)
Assigned to intensive policing	-0.095	-0.132	-0.430	-0.133
	[0.075]	[0.124]	[0.322]	[0.131]
Assigned to municipal services	-0.096	-0.010	0.200	-0.043
	[0.074]	[0.105]	[0.317]	[0.131]
<250m from any unit assigned to intensive policing	0.050	0.195	-0.043	0.144
	[0.076]	[0.140]	[0.333]	[0.139]
<250m from any unit assigned to municipal services	0.164	0.258	0.689	0.221
	[0.061]***	[0.165]	[0.275]**	[0.196]
Number of observations	1,916	414	120	294

Notes: This table reports the same intent to treat (ITT) estimates on the insecurity index as in Table 14a (Column 1) and the same analysis in three subsamples: all 414 segments assigned to Block 1 or 2 of municipal services treatment that received a city assessment (Column 2); the 120 segments in Blocks 1 and 2 that were deemed eligible for lighting improvement (Column 3); and the 294 segments that were not (Column 4). * significant at the 10 percent, ** significant at the 5 percent, *** significant at the 1 percent.

of the municipal services. The data do not offer strong support for this conclusion, however. Both services appear to have been important.

First, we see no evidence that municipal services treatment effects were concentrated in the subset of segments that were diagnosed as needing improved lights. Table 10 reports ITT effects of both interventions (without an interaction) for the full sample and on the subsample where the city diagnosed streets as eligible for lights alongside cleanup. We do this for the full sample (Column 1) and also for the subsample of 414 hot spots where the city conducted a nighttime lights needs assessment (Columns 2 to 4).[58] The coefficient on the municipal services intervention is closer to zero in the lights eligible case (Column 3) and is not statistically significantly different from zero or the cleanup only (lights ineligible) streets.

Second, we don't see systematically larger treatment effects at nighttime, when lights could have a direct effect on crime detection. We use the recorded time of a crime in police administrative data to look at nighttime versus daytime crime, and divide perceived risk questions into those that relate to nighttime and daytime risk.[59] Table 11 reports results with the interaction. In general, the coefficients on nighttime and daytime risk have the same sign and approximate magnitude, especially looking at the sum of all treatment effects in the 75 streets where both treatments (and treatment effects) were concentrated. Indeed, note that baseline risk at daytime is smaller that during the night, so a coefficient of similar size turns out to be relatively more important at daytime that nighttime. We deem this evidence as inconclusive in whether treatment effects are different at daytime vs nighttime.

[58]Recall that 201 streets were assigned to be eligible for the municipal services treatment. At the same time we selected an additional 213 for assessment, in order to be able to have baseline data on this lights needs assessment for this analysis.

[59]The city is near the equator and so 6 p.m. to 6 a.m. roughly corresponds to dusk, dark and dawn year round. Nighttime risk questions include general risk at dusk, for a young woman to walk alone after dark, and for a young man to walk alone after dark. Daytime questions include general daytime risk and risk of talking on one's smartphone.

Table 11: Impacts on insecurity by time of day

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)	Impact of proximal spillover: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Perceived risk index, z-score (+ riskier)	0.049	-0.122	-0.086	-0.084	-0.292	0.002	0.083	-0.160	-0.075
		0.259	*0.494*	*0.644*	*0.094*	*0.511*	*0.129*	***0.085***	*0.808*
Components related to daytime risk	0.019	-0.111	0.005	-0.164	-0.270	-0.022	0.091	-0.122	-0.054
		0.321	*0.736*	*0.277*	*0.139*	*0.642*	*0.103*	*0.185*	*0.723*
Components related to risk after dark	0.069	-0.122	-0.132	-0.043	-0.297	0.016	0.066	-0.179	-0.097
		0.265	*0.245*	*0.887*	***0.098***	*0.451*	*0.191*	***0.070***	*0.896*
# crimes reported to police on street segment	0.743	0.009	0.089	-0.437	-0.339	0.143	0.255	-0.272	0.125
		0.817	*0.367*	***0.043***	*0.109*	*0.315*	***0.025***	*0.196*	*0.289*
Daytime (6 a.m. – 6 p.m.)	0.472	-0.054	0.026	-0.148	-0.176	0.071	0.232	-0.246	0.058
		0.568	*0.666*	*0.337*	*0.172*	*0.578*	***0.004***	***0.084***	*0.570*
Nighttime (6 p.m. – 6 a.m.)	0.271	0.064	0.063	-0.289	-0.163	0.071	0.022	-0.026	0.067
		0.255	*0.290*	***0.005***	*0.260*	*0.273*	*0.534*	*0.895*	*0.262*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports intent to treat (ITT) estimates of the effects of the two interventions, via a WLS regression of each outcome on treatment indicators, police station (block) fixed effects, and baseline covariates (see equation 1). Column 5 reports the sum of the three treatment coefficients.

6.4 Robustness

Impacts without re-weighting and randomization inference What would we have found if we ignored different probabilities of treatment and the unusual patterns of clustering? In Appendix C.3 we estimate "naive" treatment effects ignoring IPWs and randomization inference. Direct treatment effects are slightly smaller than our main results, but the patterns remain similar. In contrast, naive spillover effects are larger and highly statistically significant. Hence failing to account for interference between units and clustering of treatment conditions leads us to underestimate the performance of hot spot policing further.

Choice of spillover regions Tables 6 and 8 above generally show some evidence of spillovers within 250 meters at the pre-specified p=0.1 level. And Appendix C.11 shows little evidence of spillovers in the 250–500 meter range for the experimental sample of streets, while for the larger sample of the non-experimental streets we see virtually no evidence of spillovers in this region. Nonetheless, an argument could be made for no spillovers.

Recall that we have 4×4 experimental conditions (in fact, 4×5 when we consider the segments ineligible for municipal services). Table 2 above reported the means. An alternative test of spillovers is to test for a statistically significant difference between pairs of columns in Table 2 (for municipal services) and then between the rows (for intensive policing). Table 12 reports the p-values from an F-test of joint significance for each of these comparisons, first comparing 250–500 meter to $>$ 500 meter segments, then $<$250 meter to $>$ 250 meter segments. We report F-tests for the officially reported crime measure alone.

We do not see evidence of statistically significant spillovers of intensive policing. None of the p-values are below 0.2. We do see some indication of proximal spillovers from municipal services in one of the two outcomes (crime incidence) and is not statistically significant in the large non-experimental sample. Thus there is a reasonable argument for calculating treatment effects ignoring spillovers. In retrospect, our pre-specified test for spillovers should have accounted for interactions between treatments, as well as baseline covariates, and addressed how we would treat economically large spillovers around or below $p = 0.1$. We report the evidence with and without spillovers here and allow readers to draw their own conclusions.

Spillovers using a continuous rate of decay Table 13 reports spillovers into the non-experimental sample of segments nearby treated hot spots, where we impose a monotonic and continuous functional form on the decay of spillovers. We ignore interactions between treatments for simplicity, as they yield similar results. The table reports estimates using an exponential rate of decay, a functional form that places some of the heaviest weight on immediately proximate streets. Linear, logarithmic, and inverse square decay functions produce qualitatively similar conclusions, even though they give more weight to more distant segments. We can interpret the coefficients in 13 as the increase in reported crimes as a street in the non-experimental sample moves a standard deviation closer to a treated hot spot.

The signs on the intensive policing coefficients are all positive but not statistically significant. This is largely consistent with the analysis of spillover regions, above. One difference is that the evidence of displacement is no longer confined to property crimes. Here the majority of displacement seems to be associated with violent crimes.

The signs on municipal services, meanwhile, are negative, implying a diffusion of benefits to nearby streets. The decrease is roughly significant at the 10% level for all crimes, and roughly significant at the 5% level for violent crimes alone. These signs are consistent across most functional forms although the statistical significance is not.

Table 12: Testing for spillovers: F-tests of weighted mean differences between control regions

Outcome	p-value from F-test of joint significance			
	Experimental sample (N = 1,919)		Non-experimental sample (N=77,848)	
	250–500m vs >500m regions	<250m vs >250m regions	250–500m vs >500m regions	<250m vs >250m regions
	(1)	(2)	(3)	(4)
A. Intensive policing				
Perceived risk	0.235	0.717		
Crime incidence	0.542	0.716		
# crimes reported to police	0.626	0.165	0.277	0.224
B. Municipal services				
Perceived risk	0.667	0.648		
Crime incidence	0.434	0.093		
# crimes reported to police	0.434	0.029	0.576	0.552

Notes: There are 4 × 7 experimental conditions, with means reported in Table 2. This table tests for mean differences iteratively, first between the >500 meter and 250–500 meter conditions, then between the <250 meter and >250 meter conditions. It does so for each intervention. For instance, to test for spillovers in the distant spillover region from from municipal services, we calculate the mean differences between the four cells in column 3 of Table 2 and the adjoining cells in column 4. This table reports the p-value from the F-test of those four mean differences.

Table 13: Spillovers to non-hot spots using a continuous exponential rate of decay, with RI p-values

	Control mean	Impact of a one standard deviation change in the average exponential distance to a hot spot treated with:	
		Intensive policing	Municipal services
	(1)	(2)	(3)
# crimes reported to police on street segment	0.274	0.049	-0.050
		0.309	*0.102*
# property crimes only	0.100	0.004	0.001
		0.788	*0.957*
# violent crimes only	0.174	0.045	-0.051
		0.303	*0.051*

Notes: Randomization inference p-values are in italics. This table estimates the coefficients on spillovers, $\check{\lambda}$, using equation 3 above. We estimate the regression on the nonexperimental sample of non-hot spots alone. The weighted distance measures have been standardized to have zero mean and unit standard deviation.

Program impacts on insecurity, ignoring spillovers Table 14 reports results from estimating equation 1 without spillovers, with and without the interaction between treatments. Conventional standard errors clustered at the quadrant level now produce reliable estimates (with tests of statistical significance nearly identical to the RI method).

The main change is that the direct effects of treatment on reducing crime are much more statistically significant. Accounting for spillovers had little effect on the estimated coefficients. They principally changed the precision of the estimates, in part because spillovers effectively reduce the sample size for the treatment-control comparison, and in part because of the spatial clustering that spillovers induce.

Looking at panel (a), without the interaction, intensive policing reduces the overall index of insecurity by about 0.12 standard deviations, and municipal services reduces it by about 0.16 standard deviations. Both perceived risk and crime incidence fall but, for intensive policing at least, the fall in crime incidence is not statistically significant. As before, we see the largest and most statistically robust impacts of state presence in the fully interacted model, particularly in the segments that received both interventions. Looking at the overall insecurity index, we estimate that policing alone or municipal services alone reduced insecurity by 0.05 and 0.07 standard deviations (not significant), but that insecurity fell 0.31 standard deviations in the 75 streets with both interventions. The effect of both interventions is huge, reducing reported crimes by more than a third.

6.5 Program impacts on state trust and legitimacy

Finally, what effect did these interventions have on trust, satisfaction with, and perceived legitimacy of the state? We pre-specified three secondary outcomes designed to capture these downstream effects of security:

1. **Opinion of police index.** This index averages the answers to four questions about the respondent's attitude towards the metropolitan police: how much trust they have in the police; how they would rate the quality of work the police do; how satisfied they are with the police; and how likely they would be to provide information to the police to improve the security of their neighborhood. Each question has a scale from 1 to 4 ranging from "a lot" to "not at all".

2. **Opinion of mayor index.** This index asks the same four questions as in the case of the police, but about the Mayor's office.

3. **Crime reporting (collaboration).** Our final measure asks people how likely they would be to report a crime to the police or other authorities, on a scale from 1 to 4 ranging from very likely to not at all likely. This is helpful in part to understand whether administrative crime reporting changes with treatment, but is also a measure of collaboration. There is a similar question in each of the police and Mayor opinion questions. In the state building and especially the counter insurgency literatures such civilian information, tips, and collaboration are among the chief indicators of state legitimacy.[60]

Table 15 reports intent to treat effects of the two interventions from estimating equation 1. We report treatment effects on the components of each index in Appendix C.9 and B.10.1.

Broadly speaking, we do not see clear evidence that either intervention increased perceived trust or legitimacy. Rather, we see an unexpected pattern: intensive policing and municipal services alone are associated with increases in the opinion of police and Mayor, but this is effectively cancelled out when

[60]We also envisioned being able to capture details of time to report, or calls to hot lines, but were unable to obtain these data.

Table 14: Impacts on insecurity, ignoring spillovers

(a) No interaction between treatments

Dependent variable	Control mean	ITT of assignment to: Intensive policing	Municipal services
	(1)	(2)	(3)
Insecurity index, z-score (+ more insecure)	0.078	-0.123	-0.160
		[.060]**	[.067]**
Perceived risk, z-score (+ riskier)	0.033	-0.116	-0.119
		[.059]**	[.065]*
Crime incidence, z-score (+ more crime)	0.096	-0.089	-0.147
		[.059]	[.068]**
Perceived & actual incidence of crime, z-score	0.039	-0.034	-0.118
		[.061]	[.072]
# crimes reported to police on street segment	1.178	-0.170	-0.164
		[.096]*	[.105]

(b) With interaction between treatments

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)
	(1)	(2)	(3)	(4)	(5)
Insecurity index, z-score (+ more insecure)	0.078	-0.049	-0.070	-0.192	-0.311
		[.055]	[.088]	[.130]	[.096]***
Perceived risk index, z-score (+ riskier)	0.033	-0.061	-0.053	-0.143	-0.257
		[.052]	[.086]	[.131]	[.096]***
Crime index, z-score (+ more crime)	0.096	-0.020	-0.065	-0.176	-0.261
		[.053]	[.089]	[.130]	[.099]***
Perceived & actual incidence of crime, z-score	0.039	-0.047	-0.134	0.033	-0.148
		[.053]	[.089]	[.139]	[.110]
# crimes reported to police on street segment	1.178	0.036	0.083	-0.530	-0.412
		[.091]	[.141]	[.202]***	[.147]***

Notes: This table reports intent to treat (ITT) estimates of the effects of the two interventions, via a WLS regression of each outcome on treatment indicators, police station (block) fixed effects, and baseline covariates (see equation **??**. Panel (a) constrains the coefficient on the interaction term to be zero, and panel (b) does not. The treatment effects in panel (b) report the marginal effect of receiving any treatment or of both, and Column 5 reports the sum of the three treatment coefficients. Standard errors are clustered using the following rules: (i) for all treated segments except with cluster size 2, each segment is a cluster; (ii) for all other untreated segments, each segment gets its own cluster identifier; (iii) for entirely untreated quadrants, they form a cluster; and (iv) for quadrants with exactly 2 units assigned to treatment, those units form a cluster. The measures of perceived risk, perceived incidence of crime, and proportion reporting crime come from our citizen survey, and the # of crimes reported to the police come from police administrative data. * significant at the 10 percent, ** significant at the 5 percent, *** significant at the 1 percent.

both treatments are received. This pattern is statistically significant when we ignore spillovers, but less robust when accounting for spillovers. The same patterns are apparent when we examine the components individually (Appendix B.10.1). In the tables ignoring any interactions (in Appendix C.9), intensive policing and municipal services are associated with little change in opinions of police, and a slightly negative effect on Mayoral opinion—a 0.13 standard deviation fall, significant at the 10% level. Overall, there is certainly no evidence of an improvement in attitudes to the police or municipal government.

7 Discussion and conclusions

Not surprisingly, we find direct state presence deters crime and violence. More surprisingly, we see some evidence of increasing returns to state presence. But probably most important of all is the divergent patterns of spillovers we observe. Most of all, we see evidence that intensive policing pushed property crime around the corner. State presence seems to have reduced violent crimes, however, and reductions on the worst forms of violence are on the margin of conventional levels for statistical significance. In spite of our large sample, there is still substantial uncertainty. Also, the spillovers estimate ignores non-spatial spillovers, and the 250 meter radius is a crude simplification. Nonetheless, some findings are clear: the small direct effect of state presence, the differential effect on property and violent crimes, non-decreasing property crimes, and a fall in murders and rapes. Thus our study is a good example of a policy evaluation where the implications hinge on how to interpret estimates and significance levels under uncertainty.

One interpretation of the violent-property crime differential is the underlying motivation of each. To the extent that property crimes are planned, state presence may simply displace calculating offenders. To the extent that violent crimes, especially homicides and assaults, are committed in the heat of the moment, aggravated by alcohol and drug abuse, state presence may deter or diffuse heated situations. We did not pre-specify this heterogeneity analysis and thus we need to interpret these results cautiously. It also seems implausible that property crime increased on net by so much, and we take this result with caution too. It is plausible that disrupting criminal activity does lead to a rise in the total number of crimes (or reported crimes) as a perverse side effect. Nonetheless the evidence certainly doesn't support a net fall in crime.

Cost-benefit considerations Were the interventions worthwhile? Given our results, this is in the eye of the beholder. On the one hand, the interventions had little marginal cost, since the city simply reallocated existing resources from some streets to others without raising their budgets or personnel.[61] While property crimes may have risen, a hundred fewer people killed or raped is enormously important. This is a trade off that many police chiefs and mayors might make.

On the other hand, reallocating street-level bureaucrats had some real costs. There was a logistical cost of coordinating police patrols to spend more time on particular streets, especially management time. The treatment assignment also made police patrols spend more time in unpleasant places, and made their jobs more difficult. In our interviews with patrol officers, most said they disliked the loss of autonomy and flexibility or the close monitoring of their movements. There are also important opportunity costs to consider. Intensive policing was a major reform, and like any bureaucracy, the police can only undertake so many reforms in a year. Because the reform required a real change in how police are monitored and managed, the Mayor's office used a fair amount of social and political capital to implement it. Thus we have

[61] If police patrols increased arrests or seizures as a result of the interventions, this would have increased costs to other agencies in the criminal justice system. But our first stage results suggest these operations did not increase.

Table 15: Impacts on state legitimacy allowing spillovers within 250 meters, with RI p-values

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)	Impact of proximal spillovers: Any intensive policing	Any municipal services	Both interventions	Sum of (6), (8), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Opinion of police, z-score (+ better)	0.024	0.143	0.210	-0.308	0.045	-0.024	0.043	0.123	0.141
		0.150	*0.107*	***0.017***	*0.867*	*0.590*	*0.817*	*0.338*	*0.797*
Opinion of mayor, z-score (+ better)	-0.014	0.001	0.179	-0.414	-0.234	-0.024	0.068	-0.025	0.020
		0.912	***0.078***	***0.003***	***0.008***	*0.523*	*0.982*	*0.919*	*0.668*
Likeliness to report crime (0-3, + higher)	2.046	0.004	0.021	0.035	0.060	-0.007	0.007	0.026	0.026
		0.921	*0.800*	*0.522*	*0.385*	*0.688*	*0.991*	*0.638*	*0.837*

Notes: This table reports intent to treat (ITT) estimates of the effects of the two interventions, via a WLS regression of each outcome on treatment indicators, police station (block) fixed effects, and baseline covariates (see equation 1. Panel (a) reports results ignoring spillovers, and panel (b) allows for spillovers within 250 meters. The treatment effects in panel (a) and panel (b) under ITT assignment report the marginal effect of receiving any treatment or of both, and Column 5 reports the sum of the three treatment coefficients in both cases. The treatment effects in panel (b) under Impact of proximal spillovers report displacement effects within 250 meters of treated hot spots. In panel (a) standard errors are clustered using the following rules: (i) for all treated segments except with cluster size 2, each segment is a cluster; (ii) for all other untreated segments, each segment gets its own cluster identifier; (iii) for entirely untreated quadrants, they form a cluster; and (iv) for quadrants with exactly 2 units assigned to treatment, those units form a cluster . In panel (b) we report randomization inference p-values. The measures of opinion of police, opinion of mayor and likeliness to report a crime come from our citizen survey.

to measure this reform against all possible other major reforms that this one supplanted.

How do our results line up with the US evidence? This experiment provides some of the first rigorous evidence on place-based crime interventions outside the U.S.[62] At first glance, it might seem that the displacement of total crime to nearby streets runs against the literature and a staple intervention in U.S. policing. Naturally, we have to compare with caution. Bogotá and the U.S. are different contexts. Policing interventions also take different forms, and vary in terms of intensity, concentration, crimes targeted, duration, and quality of approach.

That said, on close inspection, our results are not necessarily so different, since the previous literature has not ruled out positive or negative spillovers in a definitive way. There have been seven experimental studies of intensive policing that examine spillovers, with sample sizes of 24 to 120 hot spots in the experimental sample.[63] There have also been two observational studies that examine spillovers with some degree of statistical power.[64] First, studies are split on whether they observe displacement of crime or diffusion of benefits on average.[65] Second, most sample sizes are fairly small and so the confidence intervals on spillover effects are wide. When papers report confidence intervals, they typically include sizable displacement effects, even if the average points to diffusion of benefits.[66]

Methodological lessons What matters most about this Bogotá result is not whether it generalizes to the U.S. or not, or runs against the literature, but the methodological lesson for future security and other state place-based interventions. Small sample sizes will simply not help answer the crucial question of spillovers. And unless the literature can move to randomizations in multiple cities at once, conventional estimation methods simply will not work with large samples.

When small spillovers matter, anything that could bias spillover effects or make them less precise matter a great deal. This points to the importance of eliminating these biases and having accurate, efficient estimates. Failure to account for the biases arising from spillover estimation will have profound effects on our conclusions, whether it is the bias correction through IPW and re-centering, or randomization inference for calculating exact p-values. Randomization inference has yet to gain much currency in randomized trials, in part because most times they provide more or less the same conclusion as the usual clustered standard errors. Design-based

[62]Two ongoing projects in Latin America and the Caribbean are Collazos et al. (2016) in Medellín, Colombia and Sherman et al. (2014) in Trinidad and Tobago.

[63]We exclude non-patrolling studies, including one that studies police delivery of civil remedies such as orders to clean a property (Mazerolle et al., 2000). That study shows a positive diffusion effect, albeit it does not appear to be a statistically significant one (though it is difficult to say as the paper does not list standard errors or test statistics). Also, one of the papers identified by Braga et al. (2012) as showing positive diffusion effects, by Sherman and Rogan (1995), does not have data on spillovers and because of the small and fleeting treatment effects refers to displacement as a "moot point". We exclude this study.

[64]We ignore those with fewer than 15 hot spots or clusters as being too imprecise for a meaningful discussion of spillovers. Most of these studies have just 1 to 3 treatment clusters.

[65]Four of the nine studies find evidence of net displacement, four find evidence of net diffusion, and at least one is ambiguous depending on what outcome is used. Appendix **??** summarizes this classification.

[66]This can be difficult to judge, however, since several studies do not report standard errors or confidence intervals. Given that sample sizes are often under 100 or even under 30, it seems reasonable to assume that the confidence intervals include displacement effects. We are also concerned that the estimates could be biased, or the confidence intervals too small. For instance, it is unclear how imbalance along pre-intervention covariates influences results; whether catchment areas are overlapping; whether treatment and control units are close enough to lead to interference between units; and when estimates account for clustering of the treatment and of hot spots. On the latter point, for instance, Braga et al. (1999) find statistically significant direct and spillover effects with $p<.05$ or even $p<.01$, with a sample of 12 treated and 12 control units with 16 months of data. This level of significance is implausible. It is possible that statistically significant results are due to taking weekly or monthly data on the experimental units and treating these unit-months or unit-weeks as independent rather than clustering standard errors. This would dramatically overstate statistical significance. There is often insufficient information to judge, however. In such small samples, however, we cannot rely on standard distributional tests and so randomization inference or some other small-sample test of significance should be used.

estimation of spillovers, however, where units have widely different probabilities of assignment to different experimental conditions, is a textbook case for randomization inference. This problem extends to any other situation in which the structure of the clustering of experimental units in a given treatment condition is difficult to model, which is prevalent in dense networks with a high chance of outcomes or even treatments spilling over to close units.

Finally, flexibility in measuring spillovers is crucial, and we illustrate how this can be a design-based choice. in Bogotá we find evidence of spillovers in a catchment area considerably wider than the usual catchment area, which if true could mean that the aggregate effect of displacement is considerably greater. The U.S. studies typically compare crime in the two blocks (and sometimes the 150 meters) surrounding treatment and control hot spots. This two-block catchment area is probably smaller than our 250 meter radius. Many more streets fall in a 250 meter radius than a 150 meters one (in Bogotá the difference is 57,310 versus 24,987 segments close to hot spots).

Lessons for crime prevention and state building From the narrow perspective of crime and violence reduction, these results are consistent with a tenet of criminology, that crime and violence are highly concentrated in extremely specific places. But targeting coordinating and concentrating resources in the places where crime occurs may not be as effective as often believed if crime is easily displaced. In this case, it might be wiser to target the specific people who commit crimes or particular behaviors. This is the spirit of focussed deterrence, which identifies the small group of people who commit serious crimes and use threats and incentives to keep them from offending (Kennedy, 2011). This is also the spirit of cognitive behavioral therapy, which fosters skills and norms of non-violent behavior in high-risk young adults (Heller et al., 2015; Blattman et al., 2017). These may be more profitable approaches in future.

From the broader perspective of state building, this evidence shows that the effort to build the last mile of the state in Bogotá has parallels to a much broader set of cases. The tendency for people to elude the state, or simply run away, is as old as state coercion. Highly targeted state interventions may simply create the illusion of local control. It may be that state coercion and state presence have to be much more general, and much more widely spread, in order to be effective. The urban crime and violence literature has pushed theory and interventions to a more and more micro level, but to be effective, interventions might have to be more broad-based and stronger in order to keep crime from getting pushed to nearby places. The monopoly of violence is inherently broad, and order is inconsistent with an ungoverned periphery. Small-scale trials may have led us to the opposite conclusion. Larger scale investigations, which are sorely needed in the U.S. and more globally, provide more precise tests.

References

Abt, T. and C. Winship (2016). What Works in Reducing Community Violence: A Meta Review and Field Study for the Northern Triangle. Democracy International, Inc, USAID, Washingon, DC.

Apel, R. (2013). Sanctions, perceptions, and crime: Implications for criminal deterrence. *Journal of quantitative criminology 29*(1), 67–101.

Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston, and N. Singh (2017). The Efficient Deployment of Police Resources: Theory and New Evidence from a Randomized Drunk Driving Crackdown in India. *Working paper*.

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy 76*, 169–217.

Beltran, I., C. Medina, L. Pineda, C. Prieto, G. Palacio, A. Ormaza, A. Garzon, A. Tauta, S. Junca, and M. Molina (2012). Estudio sobre tendencias economicas de la delincuencia organizada. Bogota: United Nations Office on Drugs and Crime –UNODC.

Berman, E., J. H. Felter, J. N. Shapiro, and E. Troland (2013, May). Effective aid in conflict zones.

Berman, E. and A. M. Matanock (2015). The Empiricists' Insurgency. *Annual Review of Political Science*.

Berman, E., J. N. Shapiro, and J. H. Felter (2011). Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq. *Journal of Political Economy 119*(4).

Blattman, C., J. Jamison, and M. Sheridan (2017). Reducing crime and violence: Experimental evidence on cognitive behavioral therapy in Liberia. *American Economic Review 107*(4).

Bloom, H. S. (1995). Minimum Detectable Effects: A Simple Way to Report the Statistical Power of Experimental Designs. *Evaluation Review 19*(5), 547–556.

Braga, A. and B. J. Bond (2008). Policing crime and disorder hot spots: A randomized controlled trial. *Criminology 46*, 577–608.

Braga, A., A. V. Papachristos, and D. M. Hurreau (2012). An ex post factor evaluation framework for place-based police interventions. *Campbell Systematic Reviews 8*, 1–31.

Braga, A., D. Weisburd, E. Waring, L. Green Mazerolle, W. Spelman, and F. Gajewski (1999). Problem-oriented policing in violent crime places: A randomized controlled experiment. *Criminology 37*, 541–580.

Cassidy, T., G. Inglis, C. Wiysonge, and R. Matzopoulos (2014). A systematic review of the effects of poverty deconcentration and urban upgrading on youth violence. *Health and Place 26*, 78–87.

Cerdá, M., J. D. Morenoff, B. B. Hansen, K. J. Tessari Hicks, L. F. Duque, A. Restrepo, and A. V. Diez-Roux (2012). Reducing violence by transforming neighborhoods: A natural experiment in Medellin, Colombia. *American Journal of Epidemiology 175*(10), 1045–1053.

Chalfin, A. and J. McCrary (2017). Criminal deterrence: A review of the literature. *Working Paper*.

Chalfin, A. and J. McCrary (forthcoming 2017b). Are US Cities Underpoliced?: Theory and Evidence. *Review of Economics and Statistics*.

Collazos, D., E. Garcia, D. Mejia, D. Ortega, and S. Tobon (2016). Hotspots policing in a high crime environment: An experimental evaluation in Medellin. *In progress*.

Crost, B., J. H. Felter, and P. Johnston (2014). Aid Under Fire: Development Projects and Civil Conflict. *American Economic Review 104*(6), 1833–1856.

Di Tella, R. and E. Schargrodsky (2004, March). Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack. *American Economic Review 94*(1), 115–133.

Draca, M., S. Machin, and R. Witt (2011). Panic on the streets of london: Police, crime, and the july 2005 terror attacks. *The American Economic Review 101*(5), 2157–2181.

Farrington, D. P. and B. C. Welsh (2008). Effects of improved street lighting on crime: a systematic review. *Campbell Systematic Reviews* (13), 59.

Gerber, A. S. and D. P. Green (2012). *Field experiments: Design, analysis, and interpretation.* New York: WW Norton.

Groff, E. R., J. H. Ratcliffe, C. P. Haberman, E. T. Sorg, N. M. Joyce, and R. B. Taylor (2015). Does what police do at hot spots matter? The philadelphia policing tactics experiment. *Criminology 53*(1), 23–53.

Heller, S. B., A. K. Shah, J. Guryan, J. Ludwig, S. Mullainathan, and H. A. Pollack (2015). Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago. *NBER Working Paper No. 21178*.

Horvitz, D. G. and D. J. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association 47*(260), 663–685.

Kennedy, D. M. (2011). *Don't shoot: one man, a street fellowship, and the end of violence in inner-city America.* Bloomsbury Publishing USA.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Koper, C. (1995). Just enough police presence: Reducing crime and disorderly behavior by optimizing patrol time in crime hotspots. *Justice Quarterly 12*(4), 649–672.

Lipsky, M. (1969). *Street-level bureaucracy: Dilemmas of the individual in public service.* Russell Sage Foundation.

Mazerolle, L. G., J. F. Prince, and J. Roehl (2000). Civil remedies and drug control: a randomized field trial in Oakland, CA. *Evaluation Review 24*, 212–241.

Police Executive Research Forum, (2008). Violent crime in America: What we know about hot spots enforcement. Technical report, Police Executive Research Forum, Washingon, DC.

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

Scott, J. C. (2014). *The art of not being governed: An anarchist history of upland Southeast Asia.* Yale University Press.

Sherman, L., M. Buerger, and P. Gartin (1989). Beyond dial-a-cop: A randomized test of repeat call policing (recap). Washington, D.C.: Crime Control Institute.

Sherman, L. and D. P. Rogan (1995). Deterrent effects of police raids on crack houses: A randomized, controlled experiment. *Justice Quarterly 12*(4), 755–781.

Sherman, L. and D. Weisburd (1995). Does Patrol Prevent Crime? The Minneapolish Hot Spots Experiment. In *Crime Prevention in the Urban Community.* Boston: Kluwer Law and Taxation Publishers.

Sherman, L., S. Williams, A. Barak, L. R. Strang, N. Wain, M. Slothower, and A. Norton (2014). An Integrated Theory of Hot Spots Patrol Strategy: Implementing Prevention by Scaling Up and Feeding Back. *Journal of Contemporary Criminal Justice 30*(2), 95–122.

Taylor, B., C. Koper, and D. Woods (2011). A randomized controlled trial of different policing strategies at hot spots of violent crime. *Journal of Experimental Criminology 7*, 149–181.

Telep, C., R. Mitchell, and D. Weisburd (2014). How Much Time Should Police Spend at Crime Hot Spots? Answers from a Police Agency Directed Randomized Field Trial in Sacramento, California. *Justice Quarterly 31*(5), 905–933.

Vargas, A. and V. Garcia (2008). Violencia urbana, seguridad ciudadana y politicas publicas: La reduccion de la violencia en las ciudades de bogota y medellin. *Pensamiento Iberoamericano. Ejemplar dedicado a: (In)Seguridad y violencia en America Latina: un reto para la democracia 2*, 249–270.

Velasquez, C. (2010). Crimen organizado: orden divergente y vecindarios urbanos vulnerables. *Revista de Estudios Urbano Regionales 36*(108), 49–74.

Weisburd, D., A. A. Braga, E. R. Groff, and A. Wooditch (2017). Can Hot Spots Policing Reduce Crime in Urban Areas? an Agent-Based Simulation. *Criminology 55*(1), 137–173.

Weisburd, D. and C. Gill (2014). Block Randomized Trials at Places: Rethinking the Limitations of Small N Experiments. *Journal of Quantitative Criminology 30*(1), 97–112.

Weisburd, D. and L. Green (1995). Measuring Immediate Spatial Displacement: Methodological Issues and Problems. In *Crime and Place: Crime Prevention Studies*, pp. 349–359. Monsey, NY: Willow Tree Press.

Weisburd, D., D. Groff, and S. Yang (2012). *The Criminology of Place: Street Segments and Our Understanding of the Crime Problem.* New York: Oxford University Press.

Weisburd, D. and C. Telep (2016). Hot Spots Policing: What We Know and What We Need to Know. *Journal of Experimental Criminology 30*(2), 200–220.

Wilson, J. and G. Kelling (1982). Broken windows: The police and neighborhood safety. *Atlantic Monthly March*, 29–38.

Appendix for online publication

A Comparisons to other studies and interventions

A.1 Power analysis of the existing literature

The aggregate effects on crime are difficult to pinpoint because of the small size of most studies. Figure A.1 plots the systematically-reviewed studies by sample size and effect sizes, for both direct and spillover effects.[67] We calculate statistical power curves, representing the minimum effect size that we would expect to be able to detect with 80% confidence.[68] Note that even the largest studies do not exceed 50 or 100 treated hot spots, with a similarly modest number of spillover segments. The average effect size for direct hot spots treatment across the studies is 0.17 standard deviations, and 0.24 if statistically significant.[69] While covariate adjustment and blocking strategies could improve statistical power slightly, these would produce at best marginal gains in precision.

In Bogotá, the city tested two place-based security interventions on a scale large enough to identify direct treatment effects of 0.15 standard deviations, and spillovers as small as 0.02 standard deviations. We plot these in Figure A.1. For fairness in the comparison, we plot the power of our study measured also on the basis of sample size and the number of treated units.

Table A.1 summarizes the studies included in Braga et al. (2012). We also include more recent studies to complement the analysis. Power curves in figure A.1 include all randomized controlled trials in the table.

A.2 Policing levels by city

Figure A.2 presents the relationship between police personnel and population for selected cities. If we draw a line from the origin through the marker for Bogotá, corresponding to this 239 per 10,000 people, we see that almost no cities fall below Bogotá's level. Data for Colombia was reported by the Secretariat of Security

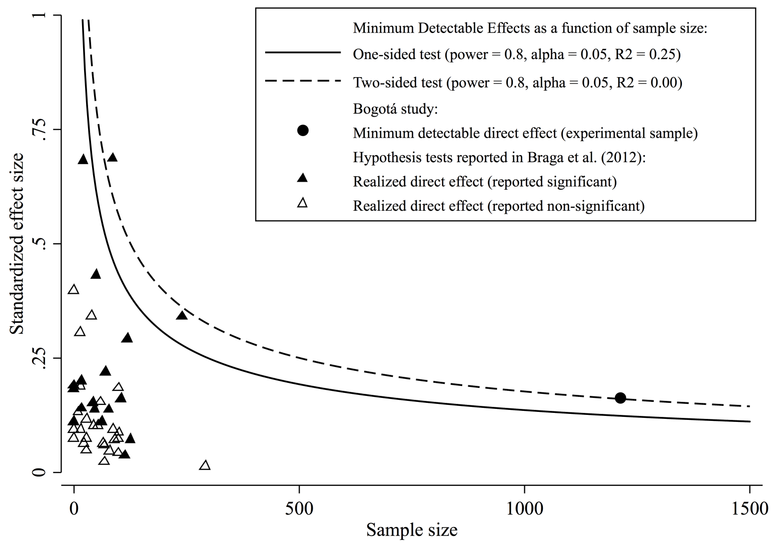
[67] The equations for the power curves are expected to be lower bounds of the actual power, as it could be increased using different randomization techniques as blocking by some specific characteristic of the units of analysis. Hence, some studies might have more power, given their sample size, than the corresponding value using the simple power formula. To make our study comparable to others, we also estimate our power using the formula rather than relying on our randomization approach. Another source of incomparability between studies could be the variation in outcomes within each experimental unit. As we present in Appendix **??**, some studies have units of analysis larger than a street segment as police beats. Some others have units of analysis smaller as specific addresses. In some cases, the main outcomes are calls for service, which might have more variation than crime reports in some contexts. Nonetheless, most of the studies focus on relatively small hot spots and we rely not only in crime reports but in an original survey of about 24,000 respondents. Hence, this source of incomparability should not be relevant.

[68] We generate the power curves assuming simple randomization and treatment assignment for half of the experimental sample. We acknowledge that some randomization procedures as blocking on pre-treatment characteristics could increase power (see for instance Gerber and Green, 2012; Weisburd and Gill, 2014), though the improvements may not be significant with small samples.

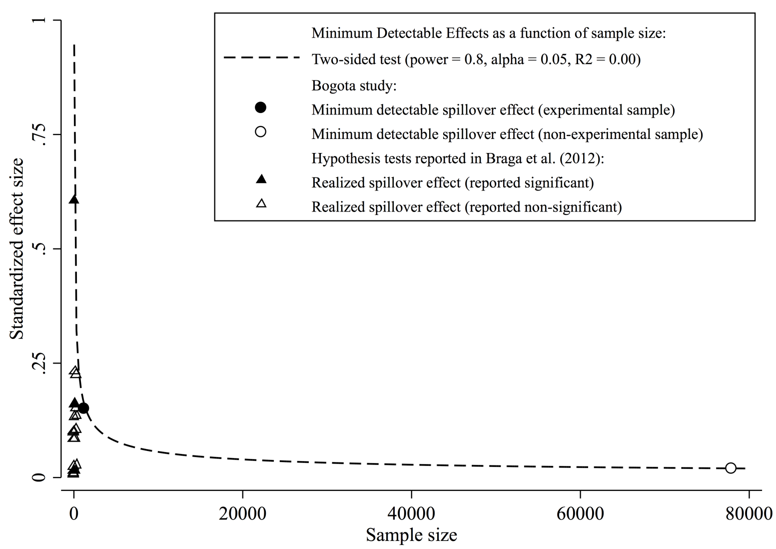
[69] We only report MDEs for studies for which it was possible to do so with the information in published papers. Generally study sizes are small, previous randomized controlled trials of intensive policing have sample sizes of 110 hot spots (55 treated) in Minneapolis (Sherman and Weisburd, 1995), 56 hot spots (28 treated) in Jersey City (Weisburd and Green, 1995), 24 hot spots (12 treated) in a different intervention in Jersey City (Braga et al., 1999), 207 hot spots (104 treated) in Kansas City (Sherman and Rogan, 1995), 100 hot spots (50 treated) in Oakland (Mazerolle et al., 2000), 34 hot spots (17 treated) in Lowell (Braga and Bond, 2008), 83 hot spots (21 treated with police patrols and 22 with problem oriented policing) in Jacksonville (Taylor et al., 2011), 120 hot spots (60 treated) in Philadelphia (Ratcliffe et al., 2011), and 42 hot spots (21 treated) in Sacramento (Telep et al., 2014). Interestingly, the first hot spots study was conducted in Minneapolis in 1989 and had a larger sample size with 250 residential addresses of which 125 were assigned to treatment and 250 commercial addressees of which also 125 were assigned to treatment Sherman et al. (1989). One of the only other large studies, by a subset of this paper's author's, is in the Colombian city of Medellín, with 384 of 967 hot spots treated (Collazos et al., 2016). Even non-experimental sample sizes have been fairly small. Di Tella and Schargrodsky (2004), for instance, examined the effects of 37 police-protected religious institutions in Buenos Aires.

Figure A.1: Statistical power in the intensive policing literature

(a) Direct and spillover effects within the experimental sample of hot spots



(b) Spillover effects into "non hot spots" proximate to the experimental sample



Notes: The figure depicts minimum detectable effects and realized effect sizes as a function of sample size. The vertical axis is in standard deviation units and measures minimum detectable effects for power curves and realize effect sizes for previous studies, and the horizontal axis measures sample size. The equations for power curves are $y = m \times 2\sqrt{\frac{1-R^2}{x}}$, where y is the standardized effect size, x is the sample size, and m is a multiple relating the standard deviation to the effect size. This multiple is 2.49 for one sided tests and 2.80 for two sided tests. See Bloom (1995) for details. Triangles represent a hypothesis test from previous studies and circles represent the minimum detectable effects in our study. See Appendix **??** for sources and a more detailed analysis of the existing literature and Appendix **??** for more details on the construction of the power curves.

Table A.1: Review of previous literature on hot spots policing

Study and reference	Main characteristics	Technical details	Spillover analysis
Minneapolis Hot spots (Minneapolis, MN). Sherman, L., & Weisburd, D. (1995). General deterrent effects of police patrol in crime hot spots: A randomized controlled trial. Justice Quarterly 12, 625- 648.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 12 months. *Main outcome:* Calls for service.	*Hot spot definition:* Hot spots are address clusters identified using the number of calls for service. *Experimental units:* 110 with 55 treated with intensive patrolling and 55 controls. *Randomization procedure:* The 110 hot spots were assigned to five blocks based on hard crime call frequencies. Then randomized treatment within each block.	No analysis on spillovers.
Jersey City DMAP (Jersey City, NJ). Weisburd, D., & Green, L. (1995). Policing drug hot Spots: The Jersey City DMAP experiment. Justice Quarterly 12, 711-36.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 15 months. *Main outcome:* Calls for service.	*Hot spot definition:* Hot spots are intersection areas identified using number of drug-related calls for service and narcotics arrests. *Experimental units:* 56 with 28 treated with intensive patrolling (focussed on drugs) and 28 controls. *Randomization procedure:* The 56 hot spots were assigned to four blocks based on call frequencies and arrests. Then randomized treatment within each block.	*Method:* Two block catchment areas surrounding treatment and control hot spots.
Kansas City Crack House Raids (Kansas City, KS). Sherman, L., & Rogan, D. (1995). Deterrent effects of police raids on crack houses: A randomized controlled experiment. Justice Quarterly 12, 755-82.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 1 day (1 raid per hot spot). *Main outcome:* Calls for service.	*Hot spot definition:* Blocks identified using calls for service and court authorized raids. *Experimental units:* 207 with 104 treated with police raids and 103 controls. *Randomization procedure:* Random assignment of treatment using the whole sample.	No analysis on spillovers.
Jersey City POP at violent places (Jersey City, NJ). Braga, A., Weisburd, D., Waring, E., Mazerolle, L.G., Spelman, W., & Gajewski, F. (1999). Problem-oriented policing in violent crime places: A randomized controlled experiment. Criminology 37, 541-80.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 16 months. *Main outcome:* Calls for service, crime reports, arrests.	*Hot spot definition:* Blocks identified using calls for service and court authorized raids. *Experimental units:* 24 with 12 treated with problem oriented policing and 12 controls. *Randomization procedure:* Hot spots were matched in couples based on qualitative and quantitative assessments. Then randomized treatment within couples.	*Method:* Two block catchment areas surrounding treatment and control hot spots. Selected hot spots were cleared so final units were separate.

Notes: Continued on following page.

Review of previous literature on hot spots policing (continued)

Study and reference	Main characteristics	Technical details	Spillover analysis
Oakland Beat Health Program (Oakland, CA). Mazerolle, L., Price, J., & Roehl, J. (2000). Civil remedies and drug control: a randomized field trial in Oakland, California. Evaluation Review, 24, 212 – 241.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 5.5 months. *Main outcome:* Calls for service.	*Hot spot definition:* Street blocks referred as having drug or blight problems. *Experimental units:* 100 with 50 treated with drug-related civil remedies and 50 controls. *Randomization procedure:* Random allocation blocking by economic use of land: residential and commercial	*Method:* 500 feet (about 150m) catchment areas surrounding treatment and control hot spots
Lowell Policing Crime and Disorder Hot Spots (Lowell, MA). Braga, A., & Bond, B. (2008). Policing crime and disorder hot spots: A randomized controlled trial. Criminology, 46 (3): 577 – 608.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 12 months. *Main outcome:* Calls for service.	*Hot spot definition:* Polygons built using spatial analysis of crime and disorder calls for service. *Experimental units:* 34 with 17 treated with problem oriented policing and 17 controls. *Randomization procedure:* The 34 hotspots were matched in couples based on qualitative and quantitative assessments. Then randomized treatment per couple.	*Method:* Two block catchment areas surrounding treatment and control hot spots. All hot spots were cleared so they included a two block catchment area to analyze spillovers.
Jacksonville Policing Violent Crime Hot Spots (Jacksonville, FL). Taylor, B., Koper, C., & Woods, D. (2011). A randomized controlled trial of different policing strategies at hot spots of violent crime. Journal of Experimental Criminology 7, 149-181.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 3 months. *Main outcome:* Calls for service and crime reports.	*Hot spot definition:* Land parcels built using spatial analysis of crime. Average hot spot size was 0.02 sq. miles. The researchers revised locations so that each hot spot was at least one block away from any other. *Experimental units:* 83 with 22 treated with problem oriented policing, 21 treated with intensive patrolling and 40 controls. *Randomization procedure:* Hot spots were arranged in four blocks according to violent crime reports. Then randomized each of the three conditions within blocks.	Method: 500 feet (about 150m) catchment areas surrounding treatment and control hot spots. All hot spots were cleared so that no hot spot was within a range of one block from another.

Notes: Continued on following page.

Review of previous literature on hot spots policing (continued)

Study and reference	Main characteristics	Technical details	Spillover analysis
Philadelphia Foot Patrol Program (Philadelphia, PA). Ratcliffe, J., Taniguchi, T., Groff, E., & Wood, J. (2011). The Philadelphia foot patrol experiment: A randomized controlled trial of police patrol effectiveness in violentcrime hot spots. Criminology 49 (3), 795-831.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 4 months. *Main outcome:* Crime reports.	*Hot spot definition:* Hot spot patrol beats identified using spatial analysis of violent crimes, validated with the Police Department. *Experimental units:* 120 with 60 treated with intensive patrolling and 60 controls. *Randomization procedure:* The 120 hot spots were ranked based on violent crime reports and matched in couples. Then randomized treatment within couples.	*Method:* Weighted displacement quotient with 2 block catchment areas.
Minneapolis RECAP (Minneapolis, MN). Sherman, L., Buerger, M., & Gartin, P. (1989). Beyond dial-a-cop: A randomized test of Repeat Call Policing (RECAP). Washington, DC: Crime Control Institute.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 12 months. *Main outcome:* Calls for service.	*Hot spot definition:* Addresses identified using the frequency of calls for service. *Experimental units:* 250 commercial units with 125 treated with problem oriented policing and 125 controls; and 250 residential units with 125 treated with problem oriented policing and 125 controls. *Randomization procedure:* Random allocation within each group of experimental units	No analysis on spillovers.
Philadelphia Policing Tactics (Philadelphia, PA). Groff, E. R., Ratcliffe, J. H., Haberman, C. P., Sorg, E. T., Joyce, N. M., & Taylor, R. B. (2015). Does what police do at hot spots matter? The Philadelphia policing tactics experiment. Criminology, 53(1), 23–53.	Randomized controlled trial. *Intervention period:* About 7 months for problem oriented policing, 3 months for foot patrols and 7 months for the offender focused intervention. *Main outcome:* Crime reports.	*Hot spot definition:* Hot spot patrol beats identified using spatial analysis of violent crimes and validated with the Police Department. Average size was 0.044 sq. miles. *Experimental units:* 27 with 20 treated with problem oriented policing and 7 controls; 27 with 20 treated with foot patrols and 20 controls; and 27 with 20 treated with offender focused interventions and 7 controls. *Randomization procedure:* Blocked by police technique suitability according to a qualitative assessment by the Police Department.	*Method:* Weighted displacement quotient with 2 block catchment areas.

Notes: Continued on following page.

Review of previous literature on hot spots policing (continued)

Study and reference	Main characteristics	Technical details	Spillover analysis
Port St. Lucie Offender Focused Intervention (Port St. Lucie, FL). Santos, R. B., & Santos, R. G. (2016). Offender-focused police intervention in residential burglary and theft from vehicle hot spots: a partially blocked randomized control trial. Journal of Experimental Criminology, 1–30.	Randomized controlled trial. *Intervention period:* 9 months. *Main outcome:* Crime reports and arrests.	*Hot spot definition:* Aggregated census blocks (to reach homogeneity in square mileage and reported crimes) identified using a qualitative assessment of neighborhoods and reported crimes. Average size was 0.60 sq. miles. *Experimental units:* 48 with 24 treated with an offender focused intervention and 24 controls. *Randomization procedure:* 3 blocks of irregular sizes grouped according to a ranking on the rate of crimes per identified offender. Half of each blocked was randomly assigned to treatment.	No analysis on spillovers.
New York Tactical Narcotics Team (New York, NY). Sviridoff, M., Sadd, S., Curtis, R., & Grinc, R. (1992). The neighborhood effects of street-level drug enforcement: tactical narcotics teams in New York. New York: Vera Institute of Justice.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 3 months. *Main outcome:* Crime reports.	*Hot spot definition:* Streets, intersections and sets of buildings. *Non-experimental units:* 2 clusters (precincts) were targeted with tactical narcotics teams (hot spots within each precinct). *Approach:* Targeted hot spots matched with similar hot spots in a different precinct.	No analysis on spillovers.
St. Louis POP in 3 Drug Areas (St. Louis, MO). Hope, T. (1994). Problem-oriented policing and drug market locations: Three case studies. Crime Prevention Studies 2, 5-32.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 9 months. *Main outcome:* Calls for service.	*Hot spot definition:* Addresses with drug sales identified. *Non-experimental units:* 3 clusters targeted with problem oriented policing. *Approach:* Hot spot addresses were compared to other addresses on the same blocks and other blocks in surrounding areas.	*Method:* Calls for service in targeted addresses compared to calls for service in addresses at the same and surrounding blocks.
Kansas City Gun Project (Kansas City, KS). Sherman, L., & Rogan, D. (1995a). Effects of gun seizures on gun violence: 'Hot spots' patrol in Kansas City. Justice Quarterly 12, 673-694.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 7 months. *Main outcome:* Crime reports.	*Hot spot definition:* Police beats of 8 by 10 blocks. *Non-experimental units:* 1 cluster targeted with intensive enforcement on possession of firearms. *Approach:* The targeted beat was matched to a control beat according to the level of reported shootings.	*Method:* Time series analysis in 7 contiguous beats.

Notes: Continued on following page.

Review of previous literature on hot spots policing (continued)

Study and reference	Main characteristics	Technical details	Spillover analysis
Beenleigh Calls for Service (Beenleigh, AUS). Criminal Justice Commission. (1998). Beenleigh calls for service project: Evaluation report. Brisbane, Queensland, AUS: Criminal Justice Commission.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 6 months. *Main outcome:* Calls for service.	*Hot spot definition:* Suburb with addresses with large number of calls for service. *Non-experimental units:* 1 cluster targeted with problem oriented policing. *Approach:* Trends in calls for service in the targeted suburb.	No analysis on spillovers.
Houston Targeted Beat Program (Houston, TX). Caeti, T. (1999). Houston's targeted beat program: A quasi-experimental test of police patrol strategies. Ph.D. diss., Sam Houston State University. Ann Arbor, MI: University Microfilms International.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 24 months. *Main outcome:* Crime reports.	*Hot spot definition:* Beats with highest reported crime. *Non-experimental units:* 3 hot spots targeted with highly visible patrols, 3 targeted with zero tolerance patrols, 1 targeted with problem oriented policing. *Approach:* Targeted beats were matched to non-contiguous beats.	*Method:* Time series analysis in contiguous beats.
Pittsburgh Police Raids (Pittsburg, PA). Cohen, J., Gorr, W., & Singh, P. (2003). Estimating intervention effects in varying risk settings: Do police raids reduce illegal drug dealing at nuisance bars? Criminology, 41 (2): 257 – 292.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 5 months. *Main outcome:* Calls for service.	*Hot spot definition:* Nuisance bar areas with 200 meters radius. *Non-experimental units:* 37 areas targeted with police raids. *Approach:* Targeted bar areas were compared to non-nuisance 40 bar areas.	No analysis on spillovers.
Buenos Aires Police after Terrorist Attack. DiTella, R., & Schargrodsky, E. 2004. Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. American Economic Review 94, 115 – 133.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 5 months. *Main outcome:* Crime reports.	*Hot spot definition:* Street blocks with Jewish centers that received increased police presence after a terrorist attack. *Non-experimental units:* 1 cluster targeted with intensive enforcement on possession of firearms. *Approach:* Targeted street blocks compared with >800 other blocks.	*Method:* One and two blocks catchment areas surrounding targeted areas.

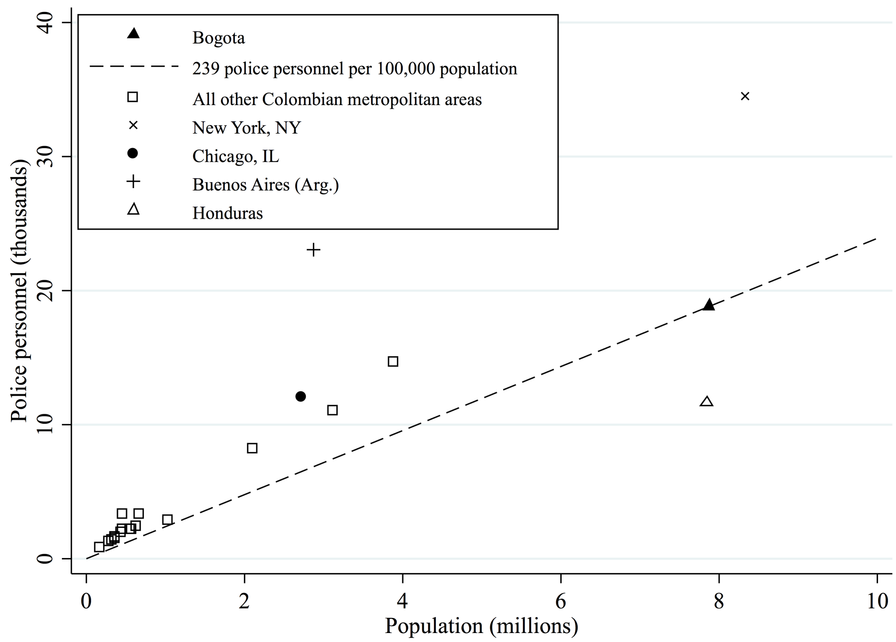
Notes: Continued on following page.

Review of previous literature on hot spots policing (continued)

Study and reference	Main characteristics	Technical details	Spillover analysis
Philadelphia Drug Corners Crackdowns (Philadelphia, PA). Lawton, B., Taylor, R., & Luongo, A. (2005). Police officers on drug corners in Philadelphia, drug crime, and violent crime: Intended, diffusion, and displacement impacts. Justice Quarterly 22, 427 – 451.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 4.5 months. *Main outcome:* Crime reports.	*Hot spot definition:* High activity drug locations with an area of 0.1 miles. *Non-experimental units:* 214 locations targeted with police crackdowns. *Approach:* Targeted locations matched with a sample of 73 other locations.	*Method:* Adjoining areas (between 0 and 0.1 miles from target sites) were compared with comparison areas (more than 0.2 miles away from target sites).
Jersey City Displacement and Diffusion Study (Jersey City, NJ). Weisburd, D., Wyckoff, L., Ready, J., Eck, J., Hinkle, J., and Gajewski, F. (2006). Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits. Criminology 44, 549 – 592.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 6 months. *Main outcome:* Crime reports.	*Hot spot definition:* Two areas comprising 81 and 88 street segments were identified according to drug sales and prostitution, respectively. *Non-experimental units:* some street segments in each area were targeted with problem oriented policing, other streets were ex-post assigned to a short range displacement area and a long range displacement area. *Approach:* Trends in prostitution and drug events were observed in targeted and displacement areas.	*Method:* Trends in catchment areas.
Boston Safe Street Program (Boston, MA). Braga, A. A., Hureau, D. M., & Papachristos, A. V. (2012). An Ex Post Facto Evaluation Framework for Place-Based Police Interventions, Evaluation Review 35(6), 592–626.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 36 months. *Main outcome:* Crime reports.	*Hot spot definition:* Streets with large number of reported violent crimes. *Non-experimental units:* 13 clusters targeted with Safe Street Program. *Approach:* Street segments within the boundaries of the targeted areas were matched to street segments outside the areas.	*Method:* Two block catchment areas surrounding targeted areas were compared to catchment areas of matched areas.

Notes: This table summarizes the studies included in Braga et al. (2012) with additional, more recent studies (noted in the main characteristics column.

Figure A.2: Police personnel and population in all Colombian metropolitan areas and other selected cities



Notes: All Colombian metropolitan areas are all with metropolitan police departments. The sources are: the Secretariat of Security of Bogota for Colombia data, the Department of Justice Statistics for U.S. data, and the United Nations Office on Drugs and Crime for other data. Data for Honduras is at the country level.

of Bogota, data for the U.S. is from the Department of Justice Statistics and other data is from the United Nations Office on Drugs and Crime. Data for Honduras is at the country level.

B Additional data and design details

B.1 Descriptive statistics and balance test for all baseline outcomes for experimental sample

Tables B.1 and B.1 expands the descriptive statistics balance table in the main paper for the full set of baseline covariates available. Columns 1–4 display the mean, standard deviation, minimum, and maximum, respectively, with each observation weighted by the inverse of the probability of being in the observed experimental condition. In columns 5–8 we perform a balance test, comparing treated units to untreated units using weighted least squares, and display both naive and randomization inference p-values.

The samples are well-balanced: only 3 out of 41 intensive policing covariates and 2 out of 41 municipal services covariates have $p_{RI} < 0.10$, or how many would be expected by chance.

B.2 Descriptive statistics for all street segments

Table B.3 displays summary statistics for all 136,982 street segments in Bogotá.

Table B.1: Descriptive statistics for the experimental sample and tests of balance for intensive policing

	Summary statistics				WLS test of balance			
	Mean	Std. Dev.	Min.	Max.	Coeff.	S.E.	p-val, ord.	p-val, RI
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
# of reported crimes, 2012-15 (original)	15.02	8.78	0	31	-0.06	0.41	0.89	0.89
# of violent crimes	5.58	4.13	0	13	0.01	0.20	0.97	0.97
# of homicides	0.37	0.66	0	2	0.04	0.03	0.30	0.32
# of assaults	5.15	3.86	0	12	-0.02	0.19	0.93	0.93
# of property crimes	9.25	6.97	0	22	-0.17	0.30	0.58	0.59
# of thefts from person	7.42	6.94	0	21	-0.04	0.29	0.89	0.89
# of car thefts	0.75	1.09	0	3	-0.07	0.05	0.18	0.22
# of motorcycle thefts	0.78	1.03	0	3	0.01	0.05	0.92	0.92
# of reported crimes, 2012-15 (updated 10/2016)	3.13	3.26	0	11	-0.19	0.16	0.23	0.24
# of violent crimes	0.55	1.01	0	4	-0.05	0.05	0.30	0.32
# of homicides	0.05	0.21	0	1	0.00	0.01	0.78	0.81
# of assaults	0.49	0.85	0	3	-0.04	0.05	0.38	0.39
# of property crimes	2.25	2.51	0	8	-0.10	0.12	0.39	0.43
# of thefts from person	1.56	1.96	0	6	-0.08	0.09	0.34	0.37
# of car thefts	0.12	0.32	0	1	-0.03	0.01	0.04	0.03
# of motorcycle thefts	0.12	0.32	0	1	0.00	0.02	0.77	0.77
# of burglaries	0.17	0.47	0	2	-0.01	0.02	0.67	0.69
# of other crimes	0.32	0.62	0	2	-0.02	0.03	0.53	0.52
# of family violence incidents	0.19	0.49	0	2	-0.02	0.02	0.50	0.51
# of sexual assault incidents	0.06	0.24	0	1	0.00	0.01	0.97	0.97
# of shoplifting incidents	0.23	0.55	0	2	-0.01	0.03	0.66	0.71
# of threats	0.09	0.29	0	1	-0.02	0.02	0.15	0.17
Average # of reported crimes in quadrant, 2012-15	3.2	3.5	0	16	0.09	0.17	0.57	0.59
Daily average patrolling time (11/2015 – 01/2016), minutes	26.3	99.9	0	3021	-1.68	4.72	0.72	0.74
Urban density	15380	20278	0	366287	-493.3	783.9	0.53	0.53
Meters from commercial center	502	550	6.8	4306	-8.25	22.50	0.71	0.77
Meters from educational center	300	242	6.2	2663	-0.89	10.61	0.93	0.94
Meters from police infrastructure	575	359	6.5	2805	-23.82	16.29	0.14	0.17
Meters from religious center	441	329	4.0	2933	-15.44	14.46	0.29	0.33
Meters from shopping center	861	711	17.4	6588	-35.99	28.83	0.21	0.26
Meters from service center	541	476	7.4	3277	-2.75	19.39	0.89	0.89
Meters from public transportation	71.5	66.6	0.01	628	2.99	3.42	0.38	0.45
Industry/commercial zone	0.36	0.48	0	1	-0.06	0.02	0.02	0.03
Services zone	0.14	0.34	0	1	0.02	0.02	0.31	0.33
High income street segment	0.09	0.28	0	1	0.00	0.01	0.78	0.78
Medium income street segment	0.56	0.50	0	1	-0.03	0.02	0.13	0.16
Municipal services assignment: Treated	0.10	0.30	0	1	-0.02	0.01	0.16	0.13
Municipal services assignment: Proximal spillover	0.27	0.44	0	1	0.06	0.02	0.01	0.01
Municipal services assignment: Distant spillover	0.27	0.44	0	1	-0.01	0.02	0.81	0.81
Municipal services assignment: Pure control	0.36	0.48	0	1	-0.04	0.02	0.11	0.20

Notes: Columns 1–4 display the summary statistics for our sample of 1,919 hotspots, weighted by the probability of being in the observed intensive policing experimental condition. In columns 5–8, we perform a balance test for treated intensive policing units vs all control units using weighted least squares. We drop segments with a zero probability of being either treated or in the control group. Column 7 displays naive p-values while column 8 displays randomization inference p-values.

Table B.2: Descriptive statistics for the experimental sample and tests of balance for municipal services

	Summary statistics				WLS test of balance			
	Mean	Std. Dev.	Min.	Max.	Coeff.	S.E.	p-val, ord.	p-val, RI
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
# of reported crimes, 2012-15 (original)	15.53	8.77	0	31	-0.44	0.54	0.41	0.43
# of violent crimes	6.02	4.25	0	13	-0.07	0.28	0.79	0.80
# of homicides	0.42	0.70	0	2	-0.01	0.05	0.89	0.91
# of assaults	5.53	3.97	0	12	-0.11	0.26	0.68	0.69
# of property crimes	9.36	7.00	0	22	-0.27	0.41	0.52	0.53
# of thefts from person	7.55	6.94	0	21	-0.35	0.38	0.36	0.37
# of car thefts	0.73	1.05	0	3	0.03	0.07	0.71	0.71
# of motorcycle thefts	0.80	1.04	0	3	0.11	0.08	0.16	0.18
# of reported crimes, 2012-15 (updated 10/2016)	3.24	3.32	0	11	-0.06	0.22	0.80	0.78
# of violent crimes	0.57	1.00	0	4	-0.02	0.07	0.78	0.80
# of homicides	0.04	0.20	0	1	-0.01	0.01	0.30	0.31
# of assaults	0.50	0.87	0	3	0.00	0.06	0.94	0.94
# of property crimes	2.33	2.55	0	8	0.04	0.17	0.80	0.78
# of thefts from person	1.63	1.96	0	6	0.05	0.12	0.69	0.68
# of car thefts	0.10	0.30	0	1	-0.01	0.02	0.64	0.66
# of motorcycle thefts	0.11	0.31	0	1	-0.01	0.02	0.81	0.81
# of burglaries	0.17	0.46	0	2	0.01	0.03	0.81	0.83
# of other crimes	0.30	0.58	0	2	-0.08	0.04	0.04	0.04
# of family violence incidents	0.17	0.45	0	2	-0.05	0.03	0.12	0.12
# of sexual assault incidents	0.05	0.23	0	1	-0.01	0.02	0.63	0.66
# of shoplifting incidents	0.24	0.55	0	2	0.02	0.04	0.61	0.61
# of threats	0.10	0.29	0	1	-0.02	0.02	0.29	0.30
Average # of reported crimes in quadrant, 2012-15	3.40	3.51	0	16	-0.01	0.20	0.94	0.94
Daily average patrolling time (11/2015 – 01/2016), minutes	1.24	0.83	0	5	0.08	0.06	0.14	0.15
Urban density	15,285	17,784	0	366287	1,019.98	1,322.01	0.44	0.47
Meters from commercial center	517	577	7	4306	15.56	33.99	0.65	0.66
Meters from educational center	297	238	6	2663	30.87	16.90	0.07	0.07
Meters from police infrastructure	549	356	6	2805	8.70	25.58	0.73	0.75
Meters from religious center	437	330	4	2933	15.14	22.46	0.50	0.51
Meters from shopping center	839	676	17	6588	0.70	45.01	0.99	0.98
Meters from service center	553	490	7	3277	24.93	30.96	0.42	0.43
Meters from public transportation	71.45	71.72	0.01	628	-1.37	5.62	0.81	0.81
Industry/commercial zone	0.39	0.49	0	1	0.06	0.04	0.12	0.12
Services zone	0.13	0.34	0	1	0.03	0.02	0.22	0.25
High income street segment	0.07	0.26	0	1	0.01	0.01	0.52	0.51
Medium income street segment	0.56	0.50	0	1	0.03	0.03	0.41	0.40
Intensive policing assignment: Treated	0.39	0.49	0	1	-0.03	0.04	0.38	0.37
Intensive policing assignment: Proximal spillover	0.37	0.48	0	1	0.00	0.03	0.91	0.89
Intensive policing assignment: Distant spillover	0.15	0.36	0	1	0.01	0.03	0.73	0.69
Intensive policing assignment: Pure control	0.09	0.28	0	1	0.02	0.02	0.39	0.45

Notes: Columns 1–4 display the summary statistics for our sample of 1,919 hotspots, weighted by the probability of being in the observed municipal services experimental condition. In columns 5–8, we perform a balance test for treated municipal services units vs all control units using weighted least squares. We drop segments with a zero probability of being either treated or in the control group. Column 7 displays naive p-values while column 8 displays randomization inference p-values.

Table B.3: Descriptive statistics for all segments in Bogota

	Summary statistics (N=136,982)			
	Mean	Std. Dev.	Min.	Max.
	(1)	(2)	(3)	(4)
# of reported crimes, 2012-15 (original)	3.84	5.08	0	31
# of violent crimes	1.73	2.44	0	13
# of homicides	0.17	0.45	0	2
# of assaults	1.55	2.25	0	12
# of property crimes	2.09	3.47	0	22
# of thefts from person	1.56	3.13	0	21
# of car thefts	0.22	0.56	0	3
# of motorcycle thefts	0.29	0.63	0	3
# of reported crimes, 2012-15 (updated 10/2016)	0.89	1.67	0	11
# of violent crimes	0.19	0.60	0	4
# of homicides	0.02	0.13	0	1
# of assaults	0.17	0.51	0	3
# of property crimes	0.56	1.22	0	8
# of thefts from person	0.36	0.91	0	6
# of car thefts	0.03	0.18	0	1
# of motorcycle thefts	0.04	0.20	0	1
# of burglaries	0.07	0.29	0	2
# of other crimes	0.15	0.42	0	2
# of family violence incidents	0.09	0.33	0	2
# of sexual assault incidents	0.03	0.16	0	1
# of shoplifting incidents	0.06	0.29	0	2
# of threats	0.04	0.19	0	1
Average # of reported crimes in quadrant, 2012-15	1.60	1.36	0	16
Daily average patrolling time (11/2015 – 01/2016), minutes	15.2	35.	0	1,387
Urban density	23,887	28,485	0	956,128
Meters from commercial center	762	688	0.3	5,114
Meters from educational center	321	239	0.6	3,834
Meters from police infrastructure	604	350	0.2	4,782
Meters from religious center	503	341	0.4	5,670
Meters from shopping center	946	650	0.2	9,181
Meters from service center	733	562	0.5	7,360
Meters from public transportation	98	76.6	0.0	2,757
Industry/commercial zone	0.21	0.41	0	1
Services zone	0.08	0.27	0	1
High income street segment	0.05	0.22	0	1
Medium income street segment	0.49	0.50	0	1
Intensive policing assignment: Treated	0.01	0.09	0	1
Intensive policing assignment: Proximal spillover	0.37	0.48	0	1
Intensive policing assignment: Distant spillover	0.34	0.47	0	1
Intensive policing assignment: Pure control	0.28	0.45	0	1
Municipal services assignment: Treated	0.00	0.07	0	1
Municipal services assignment: Proximal spillover	0.24	0.43	0	1
Municipal services assignment: Distant spillover	0.33	0.47	0	1
Municipal services assignment: Pure control	0.43	0.49	0	1

Notes: Columns 1–4 display the summary statistics for all segments in Bogota, weighted by the probability of being in the observed intensive policing experimental condition.

B.3 Additional tests of baseline balance

Our main text displays a treated versus all control comparison as our baseline balance test. Here we provide two additional balance tests. The first is treated vs all control units greater than 250m from any treated segment (thus excluding units within 250m of a treated hotspot). Tables B.4 and B.5 display these results for intensive policing and municipal services, respectively. As in the treated vs all control comparison displayed in the table, our sample is well-balanced: only 1 covariate out of 41 has $p_{RI} < 0.10$ in each of the comparisons.

The second set of balance tests, which compares proximal spillovers (units <250m from a treated segment) versus all control units greater than 250m from a treated hotspot, is displayed in Tables B.6 and B.7. Randomization was successful here, too: only 3 intensive policing and 4 municipal services covariates have $p_{RI} < 0.10$.

B.4 Endline survey sampling and subjects

In the fall of 2016, we conducted an endline survey to supplement our administrative data. We had enough funds to conduct a total of 24,000 surveys. The goal was to survey 10 individuals per segment across 2,400 segments (1919 hotspots and 480 non-experimental unit pairs that were deemed the closest matches). We aimed to survey individuals who were familiar with the street segment so we limited our sample to individuals who know, live or work in the specific segment.

Table B.8 displays a balance test for the characteristics of survey respondents for the experimental sample. The top panel displays the results at the respondent level while the bottom panel displays segment-level characteristics.We approached on average 21 individuals per segment, with a final take-up rate of 52%. Segments assigned to either treatment were no more likely to have individuals agree to the survey. Furthermore, respondents are balanced across main characteristics as only 5 out of 46 covariates have $p < 0.10$.

B.5 Clustering

In figure B.1, we display the clustering issue with our experiment. For each segment, we calculate the treatment assignment breakdown of all segments within 500m of that unit. We then take the percentage of segments that have the same treatment assignment as the initial segment as our measure. Figure B.1 displays maps of these percentages for both intensive policing and municipal services. For the hot spot policing map, we see most segments in the middle right of the map have neighbors with the same probability of assignment. This is because we have many hotspots there, and so all the non-experimental units in that section are likely to be inner spillovers. For MS, we see a ton of parts of the map where all segments have a large number of segments assigned to the same cluster. Indeed, the figure suggests that instead of having thousands of segments, we actually have 100 or so clusters that are generally assigned in the same pattern.

B.6 Inverse probability weights

Our randomization procedure gives segments variable probabilities of being in each of the treatment conditions. This is especially true for segments in our non-experimental sample. For example, non-experimental segments in relatively safer areas of Bogota have a zero percent chance of being a spillover for either treatment since there are no experimental units in those neighborhoods.

Table B.4: Test of balance for intensive policing, treated vs pooled control units

	Summary statistics		WLS test of balance			
	Mean	Std. Dev.	Coeff.	S.E.	p-val, ord.	p-val, RI
	(1)	(2)	(3)	(4)	(5)	(6)
# of reported crimes, 2012-15 (original)	15.02	8.78	1.28	0.62	0.04	0.30
# of violent crimes	5.58	4.13	0.76	0.28	0.01	0.16
# of homicides	0.371	0.657	0.09	0.04	0.02	0.15
# of assaults	5.15	3.86	0.69	0.26	0.01	0.16
# of property crimes	9.25	6.97	0.41	0.46	0.37	0.62
# of thefts from person	7.42	6.94	0.27	0.43	0.53	0.77
# of car thefts	0.753	1.091	0.02	0.07	0.72	0.77
# of motorcycle thefts	0.775	1.029	0.10	0.07	0.14	0.29
# of reported crimes, 2012-15 (updated 10/2016)	3.129	3.263	0.16	0.23	0.47	0.68
# of violent crimes	0.554	1.005	-0.01	0.08	0.88	0.91
# of homicides	0.046	0.210	-0.01	0.02	0.70	0.76
# of assaults	0.487	0.852	0.00	0.07	0.99	0.99
# of property crimes	2.25	2.51	0.12	0.17	0.48	0.69
# of thefts from person	1.56	1.96	0.05	0.13	0.72	0.84
# of car thefts	0.117	0.322	-0.04	0.02	0.12	0.19
# of motorcycle thefts	0.116	0.320	0.00	0.02	0.98	0.98
# of burglaries	0.173	0.469	0.00	0.04	0.91	0.94
# of other crimes	0.322	0.619	-0.01	0.05	0.77	0.80
# of family violence incidents	0.188	0.491	-0.02	0.04	0.53	0.57
# of sexual assault incidents	0.061	0.239	0.00	0.02	0.87	0.88
# of shoplifting incidents	0.230	0.548	0.00	0.04	0.92	0.94
# of threats	0.092	0.289	-0.01	0.02	0.73	0.76
Average # of reported crimes in quadrant, 2012-15	3.21	3.49	0.34	0.29	0.24	0.69
Daily average patrolling time (11/2015 – 01/2016), minutes	26.28	99.91	6	6	0.35	0.45
Urban density	15380	20278	-3574	2006	0.07	0.17
Meters from commercial center	502	550	-24	27	0.37	0.61
Meters from educational center	300	242	-32	18	0.07	0.33
Meters from police infrastructure	575	359	-68	25	0.01	0.13
Meters from religious center	441	329	-43	21	0.04	0.23
Meters from shopping center	861	711	-109	48	0.02	0.20
Meters from service center	541	476	-29	27	0.28	0.52
Meters from public transportation	71	67	-1	4	0.90	0.93
Industry/commercial zone	0.363	0.481	-0.032	0.037	0.39	0.52
Services zone	0.137	0.344	-0.030	0.027	0.27	0.34
High income street segment	0.089	0.284	-0.036	0.022	0.10	0.42
Medium income street segment	0.557	0.497	-0.036	0.029	0.22	0.51
Municipal services assignment: Treated	0.101	0.302	0.004	0.018	0.83	0.86
Municipal services assignment: Proximal spillover	0.267	0.442	0.113	0.032	0.00	0.04
Municipal services assignment: Distant spillover	0.270	0.444	-0.065	0.032	0.05	0.17
Municipal services assignment: Pure control	0.362	0.481	-0.052	0.029	0.07	0.28

Notes: Columns 1–2 display the summary statistics for our sample of 1,919 hotspots, weighted by the probability of being in the observed intensive policing experimental condition. In columns 3–6, we perform a balance test for treated intensive policing units vs all control units greater than 250m from a treated segment using weighted least squares. We drop segments with a zero probability of being either treated or in the control group. Column 7 displays naive p-values while column 8 displays randomization inference p-values.

Table B.5: Test of balance for municipal services, treated vs pooled control units

	Summary statistics				WLS test of balance			
	Mean	Std. Dev.	Min.	Max.	Coeff.	S.E.	p-val, ord.	p-val, RI
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
# of reported crimes, 2012-15 (original)	15.53	8.77	0	31	-0.44	0.54	0.41	0.43
# of violent crimes	6.02	4.25	0	13	-0.07	0.28	0.79	0.80
# of homicides	0.42	0.70	0	2	-0.01	0.05	0.89	0.91
# of assaults	5.53	3.97	0	12	-0.11	0.26	0.68	0.69
# of property crimes	9.36	7.00	0	22	-0.27	0.41	0.52	0.53
# of thefts from person	7.55	6.94	0	21	-0.35	0.38	0.36	0.37
# of car thefts	0.73	1.05	0	3	0.03	0.07	0.71	0.71
# of motorcycle thefts	0.80	1.04	0	3	0.11	0.08	0.16	0.18
# of reported crimes, 2012-15 (updated 10/2016)	3.24	3.32	0	11	-0.06	0.22	0.80	0.78
# of violent crimes	0.57	1.00	0	4	-0.02	0.07	0.78	0.80
# of homicides	0.04	0.20	0	1	-0.01	0.01	0.30	0.31
# of assaults	0.50	0.87	0	3	0.00	0.06	0.94	0.94
# of property crimes	2.33	2.55	0	8	0.04	0.17	0.80	0.78
# of thefts from person	1.63	1.96	0	6	0.05	0.12	0.69	0.68
# of car thefts	0.10	0.30	0	1	-0.01	0.02	0.64	0.66
# of motorcycle thefts	0.11	0.31	0	1	-0.01	0.02	0.81	0.81
# of burglaries	0.17	0.46	0	2	0.01	0.03	0.81	0.83
# of other crimes	0.30	0.58	0	2	-0.08	0.04	0.04	0.04
# of family violence incidents	0.17	0.45	0	2	-0.05	0.03	0.12	0.12
# of sexual assault incidents	0.05	0.23	0	1	-0.01	0.02	0.63	0.66
# of shoplifting incidents	0.24	0.55	0	2	0.02	0.04	0.61	0.61
# of threats	0.10	0.29	0	1	-0.02	0.02	0.29	0.30
Average # of reported crimes in quadrant, 2012-15	3.40	3.51	0	16	-0.01	0.20	0.94	0.94
Daily average patrolling time (11/2015 – 01/2016), minutes	1.24	0.83	0	5	0.08	0.06	0.14	0.15
Urban density	15,285	17,784	0	36,6287	1,019.98	1,322.01	0.44	0.47
Meters from commercial center	517	577	7	4,306	15.56	33.99	0.65	0.66
Meters from educational center	297	238	6	2,663	30.87	16.90	0.07	0.07
Meters from police infrastructure	549	356	6	2,805	8.70	25.58	0.73	0.75
Meters from religious center	437	330	4	2,933	15.14	22.46	0.50	0.51
Meters from shopping center	839	676	17	6,588	0.70	45.01	0.99	0.98
Meters from service center	553	490	7	3,277	24.93	30.96	0.42	0.43
Meters from public transportation	71.45	71.72	0.01	628	-1.37	5.62	0.81	0.81
Industry/commercial zone	0.39	0.49	0	1	0.06	0.04	0.12	0.12
Services zone	0.13	0.34	0	1	0.03	0.02	0.22	0.25
High income street segment	0.07	0.26	0	1	0.01	0.01	0.52	0.51
Medium income street segment	0.56	0.50	0	1	0.03	0.03	0.41	0.40
Intensive policing assignment: Treated	0.39	0.49	0	1	-0.03	0.04	0.38	0.37
Intensive policing assignment: Proximal spillover	0.37	0.48	0	1	0.00	0.03	0.91	0.89
Intensive policing assignment: Distant spillover	0.15	0.36	0	1	0.01	0.03	0.73	0.69
Intensive policing assignment: Pure control	0.09	0.28	0	1	0.02	0.02	0.39	0.45

Notes: Columns 1–4 display the summary statistics for our sample of 1,919 hotspots, weighted by the probability of being in the observed municipal services experimental condition. In columns 5–8, we perform a balance test for treated municipal services units vs all control units using weighted least squares. We drop segments with a zero probability of being either treated or in the control group. Column 7 displays naive p-values while column 8 displays randomization inference p-values.

Table B.6: Test of balance for intensive policing, proximal spillovers vs pooled control units

	Summary statistics		WLS test of balance			
	Mean	Std. Dev.	Coeff.	S.E.	p-val, ord.	p-val, RI
	(1)	(2)	(3)	(4)	(5)	(6)
# of reported crimes, 2012-15 (original)	15.02	8.78	1.90	0.74	0.01	0.18
# of violent crimes	5.58	4.13	1.10	0.33	0.00	0.08
# of homicides	0.37	0.66	0.05	0.04	0.24	0.48
# of assaults	5.15	3.86	1.07	0.31	0.00	0.06
# of property crimes	9.25	6.97	0.82	0.56	0.15	0.42
# of thefts from person	7.42	6.94	0.36	0.53	0.50	0.72
# of car thefts	0.75	1.09	0.17	0.08	0.03	0.07
# of motorcycle thefts	0.78	1.03	0.12	0.08	0.11	0.24
# of reported crimes, 2012-15 (updated 10/2016)	3.13	3.26	0.41	0.26	0.12	0.41
# of violent crimes	0.55	1.01	0.06	0.09	0.50	0.63
# of homicides	0.05	0.21	-0.01	0.02	0.69	0.77
# of assaults	0.49	0.85	0.06	0.08	0.39	0.54
# of property crimes	2.25	2.51	0.21	0.20	0.31	0.58
# of thefts from person	1.56	1.96	0.14	0.16	0.38	0.63
# of car thefts	0.12	0.32	-0.02	0.03	0.44	0.52
# of motorcycle thefts	0.12	0.32	0.00	0.03	0.87	0.88
# of burglaries	0.17	0.47	0.01	0.04	0.87	0.89
# of other crimes	0.32	0.62	0.02	0.06	0.74	0.79
# of family violence incidents	0.19	0.49	0.00	0.05	0.92	0.94
# of sexual assault incidents	0.06	0.24	0.01	0.02	0.72	0.74
# of shoplifting incidents	0.23	0.55	0.01	0.06	0.87	0.90
# of threats	0.09	0.29	0.01	0.03	0.71	0.75
Average # of reported crimes in quadrant, 2012-15	3.21	3.49	0.28	0.33	0.41	0.80
Daily average patrolling time (11/2015 – 01/2016), minutes	26	100	10.37	6.33	0.10	0.22
Urban density	15,380	20,278	-4,105	2,550	0.11	0.21
Meters from commercial center	502	550	10	30	0.75	0.87
Meters from educational center	300	242	-47	21	0.02	0.22
Meters from police infrastructure	575	359	-63	31	0.04	0.25
Meters from religious center	441	329	-27	24	0.26	0.54
Meters from shopping center	861	711	-117	61	0.06	0.29
Meters from service center	541	476	-47.15	32.13	0.14	0.38
Meters from public transportation	71	67	-3.00	4.63	0.52	0.66
Industry/commercial zone	0.36	0.48	0.04	0.04	0.31	0.46
Services zone	0.14	0.34	-0.08	0.03	0.01	0.04
High income street segment	0.09	0.28	-0.08	0.02	0.00	0.15
Medium income street segment	0.56	0.50	-0.03	0.03	0.35	0.62
Municipal services assignment: Treated	0.10	0.30	0.03	0.02	0.17	0.16
Municipal services assignment: Proximal spillover	0.27	0.44	0.06	0.04	0.10	0.31
Municipal services assignment: Distant spillover	0.27	0.44	-0.08	0.04	0.05	0.19
Municipal services assignment: Pure control	0.36	0.48	-0.01	0.03	0.75	0.82

Notes: Columns 1–2 display the summary statistics for our sample of 1,919 hotspots, weighted by the probability of being in the observed intensive policing experimental condition. In columns 3–6, we perform a balance test for proximal spillover intensive policing units vs all control units greater than 250m from a treated segment using weighted least squares. We drop segments with a zero probability of being either treated or in the control group. Column 7 displays naive p-values while column 8 displays randomization inference p-values.

Table B.7: Test of balance for municipal services, proximal spillover vs pooled control units

	Summary statistics		WLS test of balance			
	Mean	Std. Dev.	Coeff.	S.E.	p-val, ord.	p-val, RI
	(1)	(2)	(3)	(4)	(5)	(6)
# of reported crimes, 2012-15 (original)	15.53	8.77	-0.12	0.54	0.82	0.88
# of violent crimes	6.02	4.25	-0.51	0.26	0.05	0.18
# of homicides	0.42	0.70	-0.01	0.04	0.88	0.91
# of assaults	5.53	3.97	-0.50	0.24	0.04	0.15
# of property crimes	9.36	7.00	0.24	0.40	0.55	0.67
# of thefts from person	7.55	6.94	-0.04	0.36	0.90	0.92
# of car thefts	0.73	1.05	0.03	0.07	0.70	0.75
# of motorcycle thefts	0.80	1.04	0.13	0.07	0.05	0.12
# of reported crimes, 2012-15 (updated 10/2016)	3.24	3.32	0.19	0.20	0.34	0.49
# of violent crimes	0.57	1.00	0.09	0.06	0.16	0.22
# of homicides	0.04	0.20	0.02	0.01	0.10	0.15
# of assaults	0.50	0.87	0.06	0.05	0.25	0.31
# of property crimes	2.33	2.55	0.13	0.15	0.37	0.51
# of thefts from person	1.63	1.96	0.16	0.11	0.15	0.28
# of car thefts	0.10	0.30	0.00	0.02	0.94	0.96
# of motorcycle thefts	0.11	0.31	0.00	0.02	0.93	0.93
# of burglaries	0.17	0.46	-0.01	0.03	0.73	0.75
# of other crimes	0.30	0.58	-0.02	0.04	0.63	0.68
# of family violence incidents	0.17	0.45	-0.02	0.03	0.45	0.48
# of sexual assault incidents	0.05	0.23	0.00	0.02	0.81	0.81
# of shoplifting incidents	0.24	0.55	0.02	0.04	0.62	0.70
# of threats	0.10	0.29	0.00	0.02	0.95	0.95
Average # of reported crimes in quadrant, 2012-15	3.40	3.51	-0.44	0.19	0.02	0.30
Daily average patrolling time (11/2015 – 01/2016), minutes	1.24	0.83	-0.02	0.06	0.77	0.80
Urban density	15,285	17,784	-867.5	966.0	0.37	0.52
Meters from commercial center	517	577	13.7	26.1	0.60	0.72
Meters from educational center	297	238	35.9	14.4	0.01	0.18
Meters from police infrastructure	549	356	-11.0	20.8	0.60	0.78
Meters from religious center	437	330	-19.8	16.1	0.22	0.46
Meters from shopping center	839	676	-32.4	36.3	0.37	0.58
Meters from service center	553	490	-29.9	25.2	0.23	0.44
Meters from public transportation	71	72	-9.42	4.18	0.02	0.13
Industry/commercial zone	0.39	0.49	0.02	0.03	0.50	0.60
Services zone	0.13	0.34	0.04	0.02	0.09	0.16
High income street segment	0.07	0.26	0.04	0.01	0.01	0.36
Medium income street segment	0.56	0.50	-0.01	0.03	0.78	0.87
Intensive policing assignment: Treated	0.39	0.49	0.06	0.03	0.05	0.09
Intensive policing assignment: Proximal spillover	0.37	0.48	-0.09	0.03	0.00	0.02
Intensive policing assignment: Distant spillover	0.15	0.36	-0.01	0.02	0.55	0.67
Intensive policing assignment: Pure control	0.09	0.28	0.04	0.02	0.02	0.08

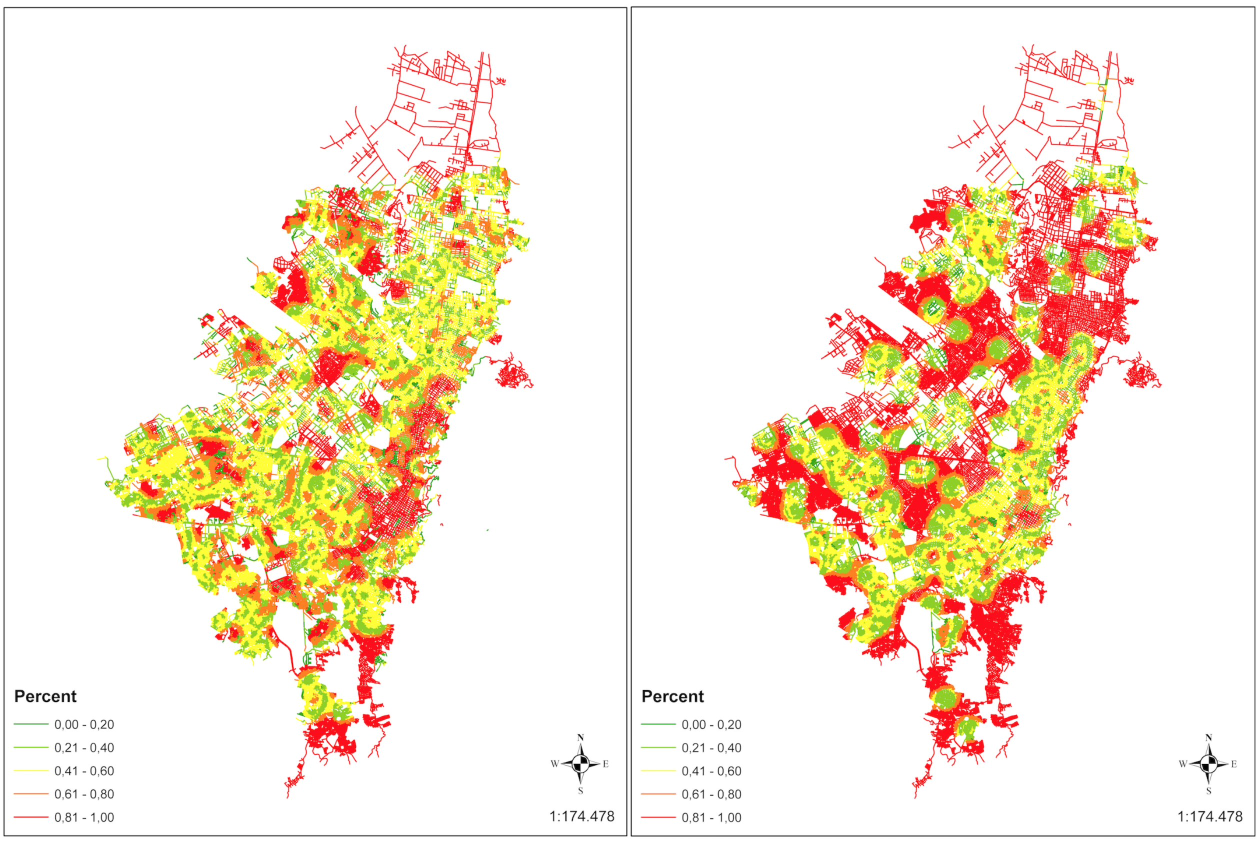
Notes: Columns 1–2 display the summary statistics for our sample of 1,919 hotspots, weighted by the probability of being in the observed municipal services experimental condition. In columns 3–6, we perform a balance test for proximal spillover municipal services units vs all control units using weighted least squares. We drop segments with a zero probability of being either treated or in the control group. Column 7 displays naive p-values while column 8 displays randomization inference p-values.

Table B.8: Characteristics of survey respondents

	Summary statistics				WLS test of balance			
					Intensive policing		Municipal services	
	Mean	Std. Dev.	Min.	Max.	Coeff.	p-val	Coeff.	p-val
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Characteristics of the respondents (respondent-level)								
Age	40.45	15.34	18.00	98.00	-0.08	0.79	0.71	0.11
Respondent is resident of the block	0.36	0.48	0.00	1.00	0.01	0.48	-0.03	0.11
Respondent work in the block	0.57	0.49	0.00	1.00	-0.01	0.35	0.02	0.24
Household size (in addition to respondent)	3.08	1.79	0.00	8.00	0.05	0.12	0.04	0.38
Total monthly income								
Less than 180.000 pesos	0.02	0.12	0.00	1.00	0.00	0.89	0.00	0.85
Between $180.000 and 450.000 pesos	0.07	0.26	0.00	1.00	0.00	0.98	0.01	0.38
Between 450.001 and 1'000.000 pesos	0.34	0.48	0.00	1.00	0.00	0.66	-0.02	0.17
Between 1'000.001 and 2'000.000 pesos	0.25	0.43	0.00	1.00	0.00	0.71	0.01	0.40
More than 2'000.001 pesos	0.11	0.32	0.00	1.00	0.00	0.93	-0.03	0.00
Don't know	0.08	0.27	0.00	1.00	0.01	0.05	0.03	0.00
Refuse to answer	0.12	0.32	0.00	1.00	-0.01	0.08	0.01	0.36
Didn't answer	0.01	0.08	0.00	1.00	0.00	0.56	0.00	0.08
Estrato	2.38	0.86	1.00	6.00	-0.03	0.13	-0.03	0.36
Highest education level								
None	0.03	0.17	0.00	1.00	0.00	0.40	0.00	0.91
Preschool	0.02	0.13	0.00	1.00	0.00	0.96	0.01	0.07
Elementary School (primaria)	0.22	0.42	0.00	1.00	0.00	0.70	0.02	0.19
High school (secundaria)	0.41	0.49	0.00	1.00	0.00	0.61	-0.01	0.40
Two-year college (técnico / tecnológica)	0.18	0.38	0.00	1.00	0.00	0.70	0.00	0.79
University	0.12	0.33	0.00	1.00	-0.01	0.24	-0.01	0.28
Graduate	0.02	0.13	0.00	1.00	0.00	0.75	0.00	0.83
Refuse to answer	0.00	0.05	0.00	1.00	0.00	0.14	0.00	0.84
Didn't answer	0.00	0.03	0.00	1.00	0.00	0.84	0.00	0.72
Female	0.48	0.50	0.00	1.00	0.01	0.13	-0.01	0.52
Charactertistics of the survey (segment-level)								
# of people approached	21.48	8.22	10.00	116.00	0.17	0.69	-0.39	0.46
# of people who rejected the survey	6.07	4.73	0.00	75.00	0.09	0.71	-0.06	0.86
# of people to busy to do the survey	3.69	4.01	0.00	46.00	0.13	0.51	-0.23	0.37
# of people who didn't qualify to do the survey	1.72	2.38	0.00	20.00	-0.07	0.55	-0.10	0.50
Take up rate	0.52	0.16	0.09	1.00	-0.01	0.43	0.00	0.80

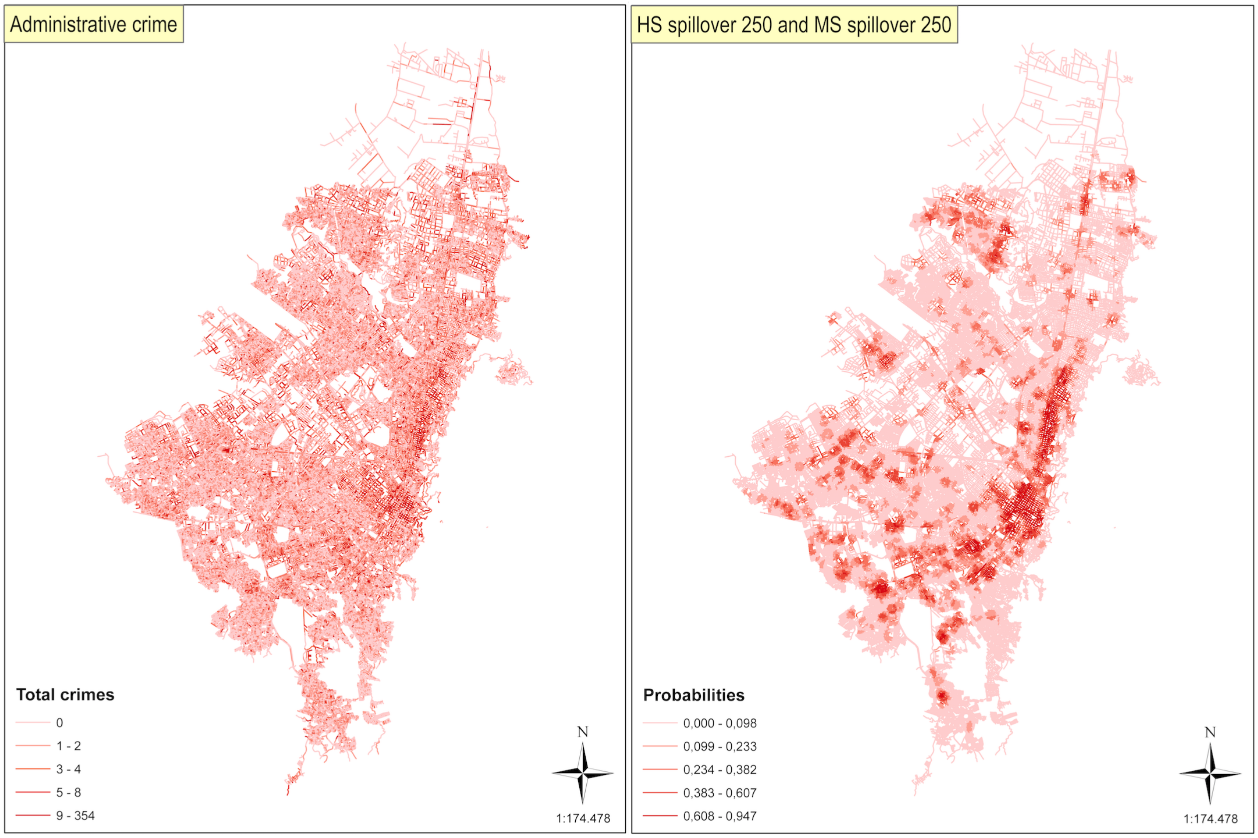
Notes: This figure displays a balance test for the endline survey. The top panel displays respondent-level characteristics such as age and gender, while the bottom panel displays segment-level characteristics such as the take up rate.

Figure B.1: Maps of percent of segments within 500m assigned to the same treatment condition



Notes: This figure displays two maps of Bogota. In the first map we display the percent of segments within 500m of each segment that are assigned to the same hotspot policing condition as that segment. In map two, we do the same thing, except for municipal services.

Figure B.2: Maps of baseline crime and probability of being proximal spillover to both interventions



Notes: This figure displays two maps of Bogota. In the first map, we display baseline administrative crime from 2012 to 2015 at the street-segment level. In the second map, we display each segment's probability of being within 250m of segments assigned to receive both interventions.

Figure B.2 compares two maps. The first map displays the number of baseline administrative crimes between 2012 and 2015 for each segment, while the second one displays each segment's probability of being within 250 meters of hotspots receiving hotspot policing and municipal services (based of 1,000 randomizations). In areas with lots of crime, non-experimental units have a higher probability of being a proximal spillover because they are located in areas with more hotspots (experimental units). In areas like the south of Bogota, however, many segments have no a zero probability of being a proximal spillover because there are no hotspots present. Thus a simple spillover vs. control comparison will lead to biased estimates on the effect of crime because the outcome (crime) is correlated with treatment assignment. In order to deal with this issue, we must use inverse probability weights and (in the case of the non-experimental units) omit units with a zero probability of being a spillover (so they are always controls) or being a control (so they are always spillovers).

B.7 Bias from inverse probability weights

In table B.9 we display the average bias associated with the use of inverse probability weights for our design. The top half shows the bias for the experimental sample while the bottom half shows the bias for the non-experimental sample. There are 1,916 units in the experimental sample, so the asymptotic requirement is unlikely to be met, leading to large biases associated with the design. By contrast, we have many more non-experimental units, which gives us much smaller biases.

B.8 Patrolling time

Figure B.3 presents the evolution of average daily patrolling time for the pre-treatment and treatment periods, as well as different groups of streets: treatment, controls (all) and non-experimental. Our estimates of average daily patrolling time are lower in the pre-treatment period because of data quality. During the pre-treatment period not all police patrols had GPS devices and some were working irregularly as the equipment was being piloted. During the treatment period there were also windows of intermittence. These malfunctioning periods, however, affected all streets equally.[70] Even though we cannot compare average daily patrolling time between the pre-treatment and treatment periods directly, the figures show that average patrolling time in control streets is between two and three times as much as that for non experimental streets. This is true for both periods and especially for time windows where the GPS devices seemed to be working better.[71]

Figure B.4 presents the distribution of quadrant $\times$ days with no GPS pings received for treatment, control and non experimental quadrants. This figure is intended to illustrate that data quality is not highly correlated with treatment. The distribution of quadrant $\times$ days with no GPS pings for both treatment and control quadrants show similar patterns. The mean number of quadrant $\times$ days with no GPS pings for treatment quadrants was 36, and it was 33 for control quadrants. Indeed, we cannot reject the null hypothesis of no difference between both means with conventional levels of statistical significance.

B.9 Tables of means

If we ignore spillovers, we can collapse our non-treated hot spots into a single control group. Table B.10 reports summary statistics for each of these 2×3 experimental conditions, using IPW for assignment into each of the treatment conditions. Starting with Panel A, we see indications that the interventions reduced the perceived risk of crime and violence. Among those eligible to receive municipal services, the policing intervention appears to reduce perceived risk. Eligible segments that did not receive municipal services saw risk decline from 0.197 to 0.141; those segments that received the cleanup saw an even larger reduction in risk, with the standardized index falling from 0.110 to -0.041. The combination of both treatments appears to have especially strong effects, reducing crime from 0.197 in the control group to -0.041 in the combined treatment group. The only segments that saw no reduction in risk were those ineligible for municipal services.

Looking next at the crime index (Panel B), we see that intensive policing is always associated with a small decline in crime. For those segments that were ineligible for municipal services, the crime index drops from -0.109 in the policing control group to -0.146 in the policing treatment group. Crime declines from 0.123 to 0.101 among those segments that were eligible for but did not receive additional municipal services and from 0.041 to -0.059 among those that did. Looking next at the effect of municipal services, we again see small improvements in crime. Among those segments that did not receive increased police attention, crime dropped from 0.123 to 0.041 with municipal services. Among those segments that did receive additional

[70] The police reported that most cases were due to software updates in all devices. For instance, to update the operating system or the software for background checks.

[71] For our estimates, we follow each GPS device chronologically, thus we track the moment at which the device enters a street and when does it leave. We made two assumptions to estimate patrolling time: (i) If we see only one GPS ping in a street and then the device moves to other streets, we impute 1 minute of patrolling time (assuming the patrol just traversed the street). (ii) If we see a device entering a street and the next ping from the same device is many hours ahead in the same street, we count until the end of the shift (assuming the device was maybe left there, but in any case the maximum patrolling time should go as much as the end of the shift).

Table B.9: IPW bias

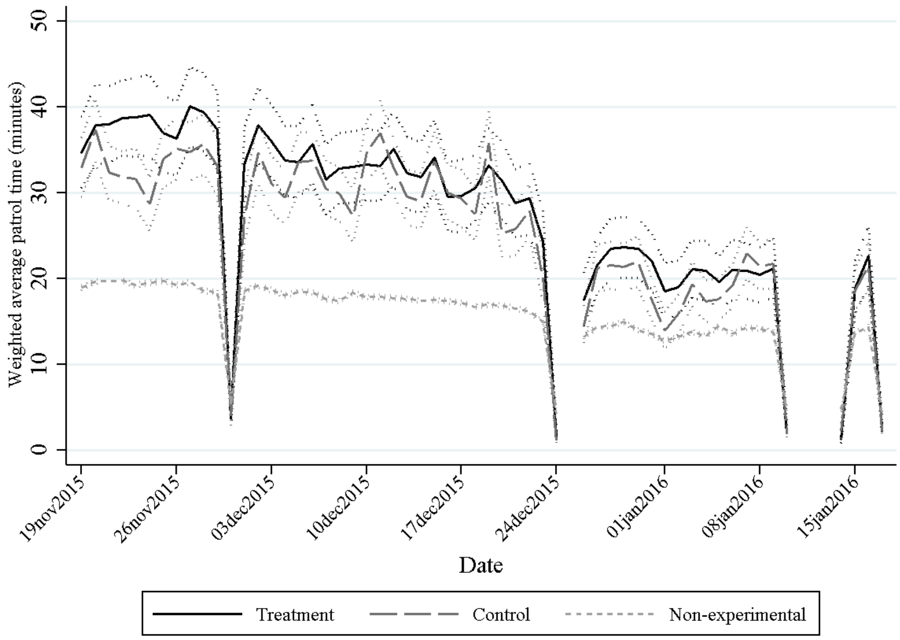
		Experimental sample							
		Treatment effect				Spillover effect			
	Interaction included?	Intensive policing	Municipal services	Interaction effect	Both (2+3+4)	Intensive policing	Municipal services	Interaction effect	Both (6+7+8)
Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Insecurity index, z-score (+ more insecure)	No	0.041	0.011		0.053	0.071	0.049		0.120
	Yes	0.035	0.003	0.022	0.060	0.081	0.060	-0.032	0.109
Perceived risk index, z-score (+ riskier)	No	0.034	0.007		0.040	0.062	0.041		0.103
	Yes	0.027	-0.003	0.023	0.048	0.072	0.051	-0.030	0.093
Crime index, z-score (+ more crime)	No	0.035	0.012		0.048	0.056	0.040		0.097
	Yes	0.032	0.007	0.014	0.052	0.064	0.048	-0.023	0.089
Perceived & actual incidence of crime, z-score (survey)	No	0.037	0.006		0.044	0.063	0.035		0.098
	Yes	0.032	-0.002	0.021	0.051	0.072	0.044	-0.027	0.089
# crimes reported to police on street segment (admin)	No	0.023	0.020		0.043	0.028	0.041		0.069
	Yes	0.023	0.022	-0.003	0.042	0.031	0.044	-0.009	0.067

		Non-experimental sample			
		Spillover effect			
	Interaction included?	Intensive policing	Municipal services	Interaction effect	Both (6+7+8)
	(1)	(6)	(7)	(8)	(9)
Insecurity index, z-score (+ more insecure)	No	0.004	0.005		0.009
	Yes	0.003	0.004	0.002	0.008
Perceived risk index, z-score (+ riskier)	No	-0.001	-0.006		-0.007
	Yes	-0.001	-0.005	-0.001	-0.007
Crime index, z-score (+ more crime)	No	0.007	0.015		0.022
	Yes	0.005	0.012	0.004	0.021
Perceived & actual incidence of crime, z-score (survey)	No	0.004	0.018		0.021
	Yes	0.001	0.015	0.004	0.021
# crimes reported to police on street segment (admin)	No	0.009	0.004		0.013
	Yes	0.010	0.005	-0.002	0.013

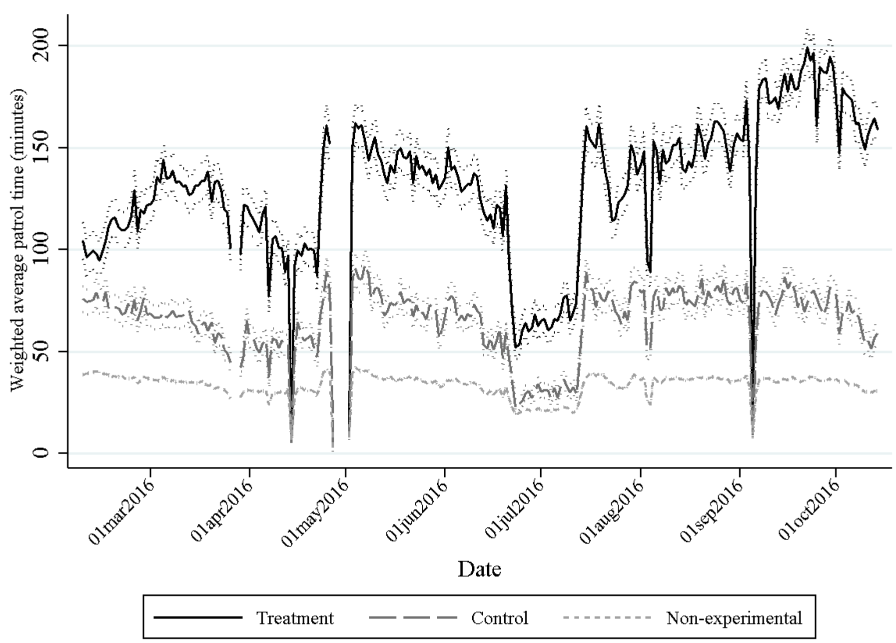
Notes: The table displays the average bias associated with the use of inverse probability weights. The first part presents the average bias for the experimental sample. The second part presents the average bias for the non-experimental sample.

Figure B.3: Evolution of patrolling time in the pre-treatment and treatment periods

(a) Pre-treatment period (November 2015 – January 2016)



(b) Treatment period (February 2016 – October 2016)



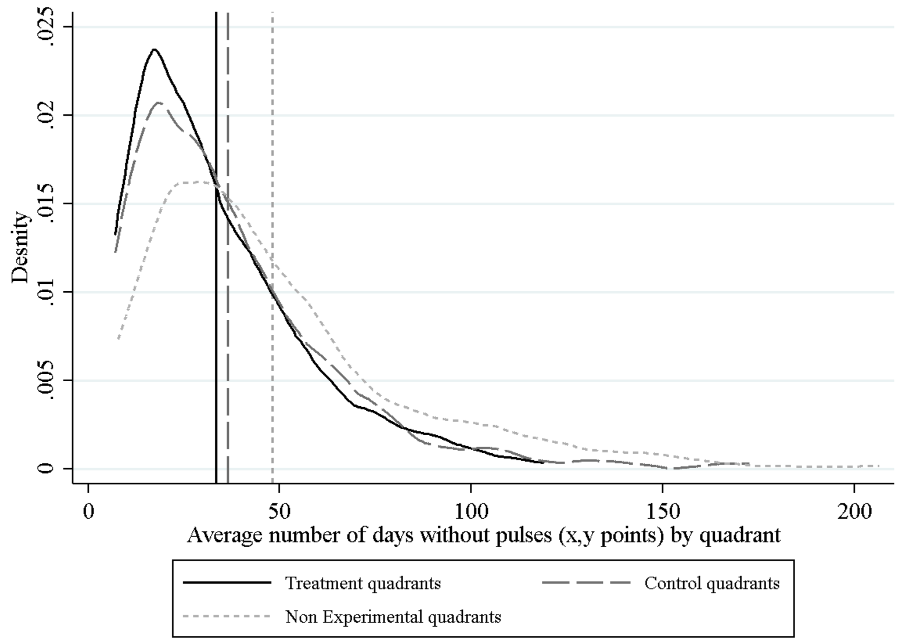
Notes: The figures present estimates of the average daily patrolling time for the pre-treatment period: November 19, 2015 through January 14, 2016, and the treatment period: February 9, 2016 through October 14, 2016. See the text in sub-section B.8 for details on the assumptions to measure patrolling time.

Table B.10: Weighted means for crime and risk, no spillovers

			Municipal services		
			Treatment	Control	Ineligible
			(1)	(2)	(3)
A: Perceived risk (z-score)					
Intensive policing	Treatment	Mean	-0.041	0.141	-0.464
		SD	0.852	1.006	0.951
		N	75	505	174
	Control	Mean	0.110	0.197	-0.495
		SD	1.061	0.959	0.978
		N	126	750	286
B: Crime incidence (z-score)					
Intensive policing	Treatment	Mean	-0.059	0.101	-0.146
		SD	0.891	0.961	0.915
		N	75	505	174
	Control	Mean	0.041	0.123	-0.109
		SD	1.010	1.085	1.023
		N	126	750	286

Notes: The table reports summary statistics for each of the 2×3 experimental conditions: treatment or control for the hot spots policing intervention, and treatment, control and not eligible for the municipal services intervention. Panel A presents weighted means for the perceived risk index and Panel B presents weighted means for the crime incidence index. Weights are the inverse of the probability of falling in the corresponding treatment status.

Figure B.4: The distribution of quadrant × days with no GPS pings received



Notes: The figure presents the distribution of the number of quadrant × days with no GPS pings received at the quadrant level. Data is presented separately for treatment, control and non experimental quadrants. The vertical lines are the mean for each group.

policing, municipal services reduced crime from 0.101 to -0.059. Notice that among those segments eligible to receive both treatments, the segments that did in fact receive the full package of interventions showed the lowest crime rate. Although only 75 segments received both treatments, the average standardized crime rate was -0.059 as compared to 0.123 among the 750 counterparts in the control condition.

B.10 Program impact ignoring spillovers: Further results

B.10.1 Impacts on all sub-components

In this subsection, we display the impacts on all sub-components of our main and secondary indices. Table B.11, while table B.12 displays the results for the opinion questions.

B.10.2 Crime type

Table B.13 divides the crime index into violent and property crime (each one is an average of survey and administrative reports of that type of crime). The patterns are broadly the same across violent and property crime, and the sums of the coefficients in column 5 are not statistically distinguishable from one another. But the impacts are proportionally much greater for violent over property crime, as there is about a third as much violent crime as property crime.

B.10.3 Impacts over time

With daily administrative crime reports, we can calculate cumulative treatment effects with each day or week of the interventions. Figure B.5 reports the results of estimating equation 1 on the cumulative level of

Table B.11: Impacts on insecurity, ignoring spillovers

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)
	(1)	(2)	(3)	(4)	(5)
Insecurity index, z-score (+ more insecure)	0.078	-0.049	-0.070	-0.199	-0.318
		[.055]	[.088]	[.130]	[.095]***
Perceived risk index, z-score (+ riskier)	0.033	-0.061	-0.052	-0.154	-0.266
		[.052]	[.086]	[.130]	[.093]***
Rank of street safety at dusk (0-3, + riskier)	1.673	-0.023	-0.012	-0.051	-0.085
		[.022]	[.037]	[.058]	[.044]*
Rank of street safety at smartphone (0-3, + riskier)	1.955	-0.021	0.008	-0.080	-0.093
		[.024]	[.041]	[.060]	[.042]**
Rank of street safety at young woman (0-3, + riskier)	2.177	-0.021	-0.047	-0.053	-0.121
		[.025]	[.039]	[.062]	[.046]***
Rank of street safety at young man (0-3, + riskier)	2.097	-0.010	-0.041	-0.066	-0.117
		[.026]	[.040]	[.062]	[.046]**
Crime index, z-score (+ more crime)	0.096	-0.020	-0.065	-0.178	-0.263
		[.053]	[.089]	[.131]	[.099]***
Perceived & actual incidence of crime, z-score (survey)	0.039	-0.047	-0.134	0.028	-0.153
		[.053]	[.090]	[.140]	[.110]
Perceived incidence of crime, z-score (survey)	0.026	-0.029	-0.127	-0.014	-0.170
		[.051]	[.090]	[.134]	[.102]*
Frequency of homicide, (0-6, + more frequent)	0.460	-0.040	-0.068	0.016	-0.093
		[.031]	[.054]	[.076]	[.054]*
Frequency of physical aggression, (0-6, + more frequent)	1.294	-0.049	-0.171	0.036	-0.184
		[.052]	[.095]*	[.141]	[.103]*
Frequency of home or business robbery, (0-6, + more frequent)	1.050	-0.015	-0.048	-0.040	-0.102
		[.048]	[.090]	[.129]	[.099]
Frequency of family violence, (0-6, + more frequent)	0.787	0.011	-0.218	0.159	-0.048
		[.042]	[.061]***	[.107]	[.088]
Frequency of whole car theft, (0-6, + more frequent)	0.640	-0.059	-0.036	-0.082	-0.176
		[.036]	[.058]	[.086]	[.066]***
Frequency of partial car theft, (0-6, + more frequent)	0.839	-0.017	-0.044	0.026	-0.035
		[.044]	[.067]	[.109]	[.084]
Frequency of whole motorcycle theft, (0-6, + more frequent)	0.607	-0.025	-0.021	-0.035	-0.080
		[.033]	[.054]	[.095]	[.077]
Frequency of partial motorcycle theft, (0-6, + more frequent)	0.565	0.001	-0.045	-0.056	-0.099
		[.036]	[.058]	[.090]	[.065]
Frequency of person robbery, (0-6, + more frequent)	2.673	-0.055	-0.109	-0.169	-0.334
		[.070]	[.116]	[.166]	[.124]***
Frequency of vandalism, (0-6, + more frequent)	2.020	-0.093	-0.208	0.177	-0.123
		[.069]	[.119]*	[.189]	[.143]

Notes: Continued on following page.

Impacts on insecurity, ignoring spillovers (continued)

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)
	(1)	(2)	(3)	(4)	(5)
Victim of any crime on the block (survey)	0.134	-0.006	-0.012	0.008	-0.011
		[.007]	[.012]	[.020]	[.015]
Victim of property crime on the block (survey)	0.124	-0.007	-0.004	0.004	-0.007
		[.007]	[.012]	[.019]	[.015]
Experienced a home robbery incident (survey)	0.010	-0.002	-0.001	-0.001	-0.005
		[.002]	[.003]	[.004]	[.003]
Experienced a business robbery incident (survey)	0.026	0.005	-0.005	0.008	0.007
		[.004]	[.005]	[.009]	[.008]
Experienced a person robbery incident (survey)	0.083	-0.011	-0.002	0.000	-0.013
		[.005]**	[.010]	[.015]	[.011]
Experienced a partial motorcycle theft incident (survey)	0.002	0.000	-0.001	0.002	0.002
		[.001]	[.001]	[.002]	[.002]
Experienced a whole motorcycle theft incident (survey)	0.002	0.000	0.003	-0.004	-0.001
		[.001]	[.002]	[.002]*	[.001]
Experienced a partial car theft incident (survey)	0.014	0.000	0.000	0.001	0.000
		[.002]	[.003]	[.006]	[.005]
Experienced a whole car theft incident (survey)	0.004	-0.002	-0.001	0.004	0.000
		[.001]***	[.002]	[.003]	[.002]
Experienced an extortion incident (survey)	0.002	0.002	0.001	-0.005	-0.002
		[.001]	[.002]	[.002]**	[.001]
Victim of violent crime on the block (survey)	0.018	0.000	-0.008	0.004	-0.004
		[.003]	[.004]**	[.006]	[.005]
Experienced a physical aggression incident (survey)	0.016	-0.002	-0.006	0.006	-0.002
		[.002]	[.003]*	[.006]	[.004]
Experienced a homicide incident (survey)	0.000	0.001	0.001	-0.002	0.000
		[0000]**	[.001]	[.001]**	[0000]
Experienced a family violence incident (survey)	0.003	0.001	-0.002	-0.002	-0.003
		[.001]	[.002]	[.002]	[.001]**
# crimes reported to police on street segment (admin)	1.178	0.036	0.082	-0.525	-0.407
		[.091]	[.142]	[.203]***	[.147]***
# violent crimes reported to police on street segment (admin)	0.264	-0.009	0.011	-0.145	-0.143
		[.033]	[.052]	[.068]**	[.055]***
# homicides during intervention (admin)	0.013	-0.005	0.002	-0.011	-0.015
		[.005]	[.010]	[.011]	[.004]***
# assaults during intervention (admin)	0.163	-0.017	-0.002	-0.066	-0.085
		[.025]	[.040]	[.052]	[.041]**
# family violence during intervention (admin)	0.043	0.007	0.021	-0.044	-0.016
		[.014]	[.024]	[.031]	[.024]

Notes: Continued on following page.

Impacts on insecurity, ignoring spillovers (continued)

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)
	(1)	(2)	(3)	(4)	(5)
# threats during intervention (admin)	0.037	0.008	-0.010	-0.017	-0.020
		[.014]	[.019]	[.026]	[.016]
# sexual assaults during intervention (admin)	0.009	-0.001	-0.001	-0.006	-0.008
		[.004]	[.006]	[.008]	[.003]***
# property crimes reported to police on street segment (admin)	0.914	0.045	0.072	-0.380	-0.264
		[.076]	[.120]	[.174]**	[.123]**
# thefts from person during intervention (admin)	0.681	0.049	-0.003	-0.179	-0.133
		[.069]	[.100]	[.149]	[.105]
# car thefts during intervention (admin)	0.055	0.007	-0.012	-0.046	-0.052
		[.014]	[.018]	[.025]*	[.016]***
# motorcycle thefts during intervention (admin)	0.056	-0.008	0.048	-0.086	-0.046
		[.014]	[.032]	[.040]**	[.016]***
# burglary during intervention (admin)	0.045	0.003	0.015	-0.039	-0.021
		[.013]	[.020]	[.028]	[.018]
# shoplift during intervention (admin)	0.077	-0.005	0.024	-0.030	-0.011
		[.015]	[.034]	[.045]	[.031]

Notes: This table reports intent to treat (ITT) estimates of the effects of the two interventions, via a WLS regression of each outcome on treatment indicators, police station (block) fixed effects, and baseline covariates (see equation **??**. The treatment effects report the marginal effect of receiving any treatment or of both, and Column 5 reports the sum of the three treatment coefficients. Standard errors are clustered using the following rules: (i) for all treated segments except with cluster size 2, each segment is a cluster; (ii) for all other untreated segments, each segment gets its own cluster identifier; (iii) for entirely untreated quadrants, they form a cluster; and (iv) for quadrants with exactly 2 units assigned to treatment, those units form a cluster . The measures of perceived risk, perceived incidence of crime, and proportion reporting crime come from our citizen survey, and the # of crimes reported to the police come from police administrative data.

Table B.12: Impacts on opinions

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)
	(1)	(2)	(3)	(4)	(5)
Opinion of police, z-score (+ better opinion)	-0.075	0.130	0.177	-0.295	0.012
		[.054]**	[.087]**	[.133]**	[.102]
Level of police trust, 0-3 (+ more trust)	1.154	0.038	0.067	-0.122	-0.017
		[.020]*	[.032]**	[.049]**	[.038]
Rating of police work, 0-3 (+ better)	1.306	0.058	0.062	-0.095	0.026
		[.017]***	[.027]**	[.042]**	[.032]
Likeliness to aid police, 0-3 (+ more likely)	1.636	0.011	0.031	-0.057	-0.015
		[.023]	[.039]	[.063]	[.048]
Level of police satisfaction, 0-3 (+ more satisfaction)	1.114	0.037	0.039	-0.061	0.015
		[.018]**	[.030]	[.046]	[.034]
Opinion of mayor, z-score (+ better opinion)	-0.056	0.069	0.217	-0.497	-0.211
		[.052]	[.103]**	[.152]***	[.108]*
Level of mayor trust, 0-3 (+ more trust)	0.924	0.020	0.066	-0.204	-0.118
		[.022]	[.037]*	[.057]***	[.042]***
Rating of mayor's office, 0-3 (+ better)	1.151	0.025	0.085	-0.172	-0.062
		[.018]	[.029]***	[.048]***	[.038]
Likeliness to aid mayor's office, 0-3 (+ more likely)	1.413	0.023	0.040	-0.096	-0.033
		[.022]	[.045]	[.068]	[.048]
Level of mayor satisfaction, 0-3 (+ more satisfaction)	0.946	0.010	0.058	-0.092	-0.024
		[.017]	[.032]*	[.050]*	[.038]

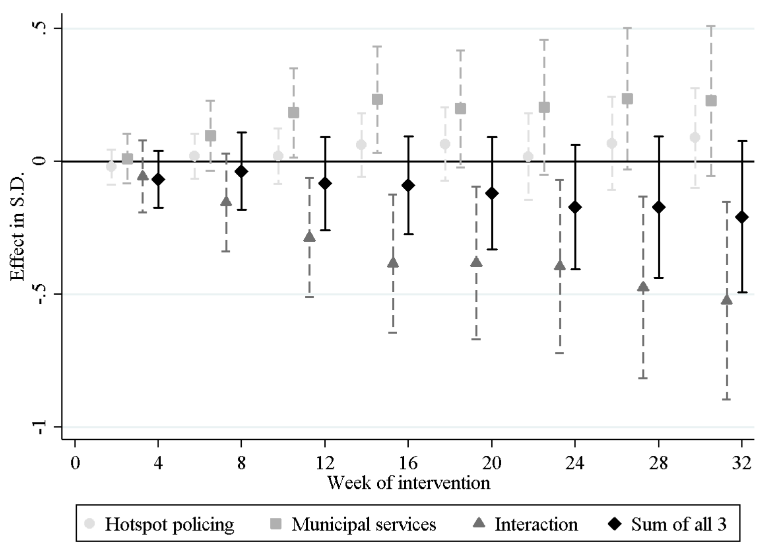
Notes: This table reports intent to treat (ITT) estimates of the effects of the two interventions, via a WLS regression of each outcome on treatment indicators, police station (block) fixed effects, and baseline covariates (see equation **??**. The treatment effects report the marginal effect of receiving any treatment or of both, and Column 5 reports the sum of the three treatment coefficients. Standard errors are clustered using the following rules: (i) for all treated segments except with cluster size 2, each segment is a cluster; (ii) for all other untreated segments, each segment gets its own cluster identifier; (iii) for entirely untreated quadrants, they form a cluster; and (iv) for quadrants with exactly 2 units assigned to treatment, those units form a cluster. The measures of perceived risk, perceived incidence of crime, and proportion reporting crime come from our citizen survey, and the # of crimes reported to the police come from police administrative data.

Table B.13: Impacts on different types of crime

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both inter-ventions	Sum of (1), (2), and (3)
	(1)	(2)	(3)	(4)	(5)
Violent crime incidence, z-score	0.069	-0.019	-0.176	-0.114	-0.308
		[.055]	[.089]**	[.128]	[.096]***
Property crime incidence, z-score	0.101	-0.009	0.002	-0.212	-0.218
		[.057]	[.096]	[.144]	[.104]**

Notes: This table reports intent to treat (ITT) estimates of the effects of the two interventions, via a WLS regression of each outcome on treatment indicators, police station (block) fixed effects, and baseline covariates (see equation **??**. Column 5 reports the sum of the three treatment coefficients. Standard errors are clustered using the following rules: (i) for all treated segments except with cluster size 2, each segment is a cluster; (ii) for all other untreated segments, each segment gets its own cluster identifier; (iii) for entirely untreated quadrants, they form a cluster; and (iv) for quadrants with exactly 2 units assigned to treatment, those units form a cluster.

Figure B.5: Impacts on reported crime over time, in weeks since treatment began (administrative crime data only)



Notes: The figure reports the ITT effect of the two interventions and the interaction term, plus the sum of these three coefficients. The sample accumulates the number of weeks of administrative data on crime reports included, starting five weeks after the intensive policing treatment began.

reported crime starting 5 weeks after intensive policing began, and continuing to the end of the intervention period. The cumulative effect of both interventions appears within 8 to 12 weeks of the intervention, and grows over time. In particular, we see that the marginal effect of receiving both interventions grows larger as time passes.

B.10.4 Could systematic measurement error drive these results?

The results in Table B.13 are driven primarily by two types of measures: survey-based risk perceptions, and administrative crime. There are no discernible impacts on survey-based crime incidence. One possible explanation is that crime is rare, and surveying just 10 people from a street not only gives us a noisy estimate of crime, but systematically underestimates crime across all treatment groups. Another possibility is reporting bias correlated with treatment, particularly the risk that police presence increases the official reporting of crimes. There is also a risk that treatment affects survey-based reporting of crime, either because it generates selection into who is out on the street during our survey, or because it influences their responses. Such misreporting appears to unlikely for two main reasons. First, we see no treatment effect on crime reporting. The survey asked respondents their likelihood of reporting a future crime to the police, on a scale of 0 to 3. The average response in control segments was 2.0, with a treatment effect [standard error] of 0.016 [.029] from policing and 0.035 [.032] from municipal services. any misreporting would have to match the pattern of our results, and be largest when receiving both policing and services but not either intervention alone. This is possible but narrows the range of plausible types of measurement.

C Additional analysis

C.1 First stage analysis

In Table C.1, we display first stage results assuming proximal spillovers. The increases in patrolling time are similar to those displayed in Table 4where we assume no spillovers. We see that patrolling time rises between about 4 and 8 minutes per day in nearby hotspots, though this is not statistically significant.

C.2 Pre-specified pairwise results

The estimation procedure used in this paper is different from the ones we described in our pre-analysis plan. In this section, we document the reasons why it was appropriate to switch estimation strategies.

Our pre-specified estimation strategy (see page 17 of the PAP) would use pairwise regressions to estimate the direct and spillover effects of the intervention. Let us assume we wanted to estimate the effects of the hot spot policing treatment given one level of spillovers, so our possible experimental conditions are: treated by hotspot policing T_H, <250m of a unit treated with hotspot policing S_H, and >250m away from a unit receiving hotspot policing (C_H, the control group). Our PAP says we would run the following WLS regression:

$$Y_{sqp} = \beta_0 + \theta_H * T_H + \beta * X_{sqp} + \gamma_p + \varepsilon_{sqp} \tag{4}$$

Our weights are determined by the probability of being either in T_H, S_H, or C_H (for example, if a street is in S_H, its weight is $\frac{1}{\Pr(S_H)}$). Furthermore, we restrict the regressions to (i) segments only in T_H or C_H, and (ii) segments with a non-zero probabilities of being in T_H and C_H (i.e. $0 < \Pr(T_H) < 1 \bigcup 0 < \Pr(C_H) < 1$). The

Table C.1: First stage analysis assuming proximal spillovers

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	ITT of assignment to: Any municipal services	ITT of assignment to: Sum of (1) and (2)	Impact of proximal spillovers: Any intensive policing	Impact of proximal spillovers: Any municipal services	Impact of proximal spillovers: Sum of (5) and (6)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Hotspot policing compliance							
Survey: Believes police presence increased in last 6 months	0.101	0.080	0.019	0.099	0.009	0.006	0.015
		0.000	*0.187*	***0.000***	*0.265*	*0.501*	*0.221*
Daily average patrolling time, excluding quadrant-days without data	78.137	68.844	19.333	88.177	8.620	4.834	13.454
		0.000	*0.113*	***0.000***	*0.339*	*0.695*	*0.287*
Excluding days below 10th and above 90th %ile	61.815	40.775	4.187	44.962	4.400	3.890	8.290
		0.000	*0.215*	***0.000***	*0.183*	*0.450*	*0.115*
# of arrests	0.149	-0.033	-0.005	-0.038	0.015	-0.022	-0.007
		0.774	*0.911*	*0.768*	*0.947*	*0.612*	*0.668*
# of drug seizure cases	0.025	0.006	0.033	0.039	0.010	-0.001	0.009
		0.852	*0.130*	*0.224*	*0.760*	*0.942*	*0.868*
# of gun seizure cases	0.005	0.007	0.006	0.013	0.004	-0.004	0.001
		0.278	*0.464*	*0.187*	*0.467*	*0.501*	*0.907*
# of recovered car cases	0.004	0.000	-0.002	-0.002	0.001	-0.001	0.000
		0.861	*0.634*	*0.700*	*0.713*	*0.592*	*1.000*
# of recovered motorbike cases	0.000	-0.016	0.032	0.016	0.005	-0.005	0.000
		0.160	***0.087***	*0.294*	*0.617*	*0.748*	*0.869*
Panel B: Municipal services compliance							
Survey: Believes mayor presence increased in last 6 mon	0.142	0.008	0.015	0.023	0.009	-0.005	0.004
		0.413	*0.248*	*0.149*	*0.410*	*0.660*	*0.706*
Eligible for lights intervention	0.045	-0.002	0.158	0.155	0.006	-0.021	-0.015
		0.810	***0.000***	***0.000***	*0.831*	*0.149*	*0.389*
Received lights intervention	0.000	0.006	0.201	0.207	0.020	0.001	0.021
		0.786	***0.000***	***0.000***	*0.230*	*0.855*	*0.421*
Eligible for garbage intervention	0.000	-0.023	0.631	0.608	-0.045	0.010	-0.034
		0.181	***0.000***	***0.000***	*0.017*	*0.731*	*0.106*
Received garbage intervention	0.000	-0.010	0.372	0.362	-0.036	0.005	-0.031
		0.395	***0.000***	***0.000***	***0.015***	*0.894*	***0.065***
June 2016 assessment							
Graffiti on block	0.712	-0.033	0.106	0.073	0.004	0.127	0.132
		0.776	*0.107*	*0.364*	*0.693*	***0.026***	***0.080***
Rubbish on block	0.117	0.065	0.053	0.118	-0.047	0.043	-0.004
		0.285	*0.442*	*0.193*	*0.887*	*0.371*	*0.657*
Broken street light on block	0.000	0.009	0.005	0.014	-0.002	0.001	-0.001
		0.177	*0.217*	*0.177*	*0.848*	*0.957*	*0.644*

Notes: Continued on following page.

Table: First stage analysis assuming proximal spillovers (continued)

Dependent variable	Control mean	ITT of assignment to:			Impact of proximal spillovers:		
		Any intensive policing	Any municipal services	Sum of (1) and (2)	Any intensive policing	Any municipal services	Sum of (5) and (6)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
December 2016 assessment							
Graffiti on block	0.619	0.001	0.072	0.072	-0.028	0.177	0.149
		0.947	*0.323*	*0.477*	*0.686*	***0.008***	*0.137*
Rubbish on block	0.238	-0.054	-0.002	-0.056	-0.121	0.024	-0.097
		0.626	*0.956*	*0.767*	*0.205*	*0.553*	*0.560*
Broken street light on block	0.018	0.025	-0.004	0.021	0.019	0.018	0.036
		0.286	*0.952*	*0.497*	*0.568*	*0.490*	*0.355*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports intent to treat (ITT) estimates of equation 1, estimating the direct effects of the two interventions (Columns 2 to 3) and the spillover effects (Columns 5 to 7) via a WLS regression of each outcome on treatment indicators, spillover indicators, police station (block) fixed effects, and baseline covariates. Columns 4 and 7 report the sum of the two preceding coefficients. Outcomes with $p < .10$ are bolded.

Table C.2: Distribution of assignments, by treatment

Panel A: Distribution of municipal service assignments

			Municipal services assignment		
		Total	Treated	<250m	>250m
		(1)	(2)	(3)	(4)
Intensive	Treated	756	0.10	0.26	0.64
policing	<250m	705	0.10	0.40	0.50
assignment	>250m	458	0.12	0.15	0.73
		1919			

Panel B: Distribution of policing assignments

			Hotspot policing assignment		
		Total	Treated	<250m	>250m
		(1)	(2)	(3)	(4)
Municipal	Treated	201	0.37	0.37	0.26
services	<250m	546	0.36	0.51	0.13
assignment	>250m	1172	0.41	0.30	0.29
		1919			

Notes: This table displays the distribution of treatment assignments for each intervention. Panel A depicts the proportion of streets assigned to the different treatment status on municipal services, within each treatment block for hot spot policing. Panel B depicts the proportion of streets assigned to the different treatment status on hot spots policing, within each treatment block for municipal services.

coefficient of interest is θ_H, which represents the ITT estimate of receiving the hot spot policing treatment on outcome Y relative to segments greater than 250m away from any treated hotspot.

This pairwise regression is incorrect because it fails to recognize the complexity of our design. We test both hot spot policing and municipal services in a factorial design, so probability weights need to be determined by the *joint* probability of hot spot policing and municipal service assignment, not just assignment to one of the treatments. Failure to account for the joint probability can mix up effects between each of the interventions. For example, if segments treated by hot spot policing have a higher chance than hot spot policing control segments to be inner spillovers for municipal services, then θ_H in equation 4 will conflate the direct effect of hot spot policing and the spillover effect of municipal services.

This is exactly what we see in our design. In table C.2, we show the distribution of treatment assignments for each intervention. Panel A shows that while segments in each hot spot policing block all have a similar proportion (~11%) of their segments receiving municipal services, segments treated with policing are more likely than segments >250m from treated policing segments to be spillover units for municipal services. In the case that there are spillover effects from municipal services, it will not be possible to use the pairwise regression detailed above to estimate just the effect of hot spot policing.

There are two changes we can make to the regressions outlined in the PAP so that our empirical strategy is compatible with the realities of our factorial design. First, we can base our probability weights off the joint probability of assignment. Second, we can insert dummies for municipal service assignment into equation 4. Making these changes gives us the following regression:

$$Y_{sqp} = \beta_0 + \theta_H * T_H + \theta_M * T_M + \theta_H * S^M + \beta * X_{sqp} + \gamma_p + \varepsilon_{sqp} \quad (5)$$

Table C.3: Hot spots policing impacts on insecurity, pre-specified regressions

	ITT of assignment to:				
	Accounting for distant spillovers	Accounting for proximal spillovers			No spillovers
Dependent variable	HSP outer spillover (1)	HSP treated (2)	HSP inner spillover (experimental) (3)	HSP inner spillover (non-experimental) (4)	HSP treated (5)
Insecurity index, z-score (+ more insecure)	0.104	-0.105	-0.056	0.119	-0.063
	0.294	*0.286*	*0.765*	*0.229*	*0.192*
Perceived risk index, z-score (+ riskier)	0.015	-0.109	-0.082	0.120	-0.067
	0.689	*0.239*	*0.477*	*0.313*	*0.148*
Crime index, z-score (+ more crime)	0.158	-0.066	-0.012	0.079	-0.038
	0.138	*0.529*	*0.912*	*0.321*	*0.443*
Perceived & actual incidence of crime, z-score	0.182	-0.084	-0.031	0.106	-0.046
	0.124	*0.369*	*0.868*	*0.338*	*0.342*
# crimes reported to police on street segment	0.073	-0.014	0.028	0.014	-0.015
	0.590	*0.970*	*0.764*	***0.082***	*0.874*

Notes: This table reports intent to treat (ITT) estimates of the effects of hotspot policing using the pre-specified regressions. Randomization inference p-values are italicized.

Including an additional indicator for being a hotspot policing spillover in this regression allows us to estimate all four effects (direct effect of hotspot policing, direct effect of municipal services, spillover effect of hotspot policing, spillover effect of municipal services) in one regression. This corresponds to the constrained version of equation **??** in the main paper where $\beta_3 = 0$. Thus the regressions used in this paper correctly estimates the effects of our factorial design by using the correct inverse probability weights and estimating all the effects in the same regression.

Nevertheless, we display the pairwise regressions pre-specified for clarity purposes. Table C.3 displays the hotspot policing effect while table displays the municipal services effects. Meanwhile, table C.5 displays the interaction effects. Most of the differences for the treatment effects are coming from the use of different weights. In table C.5 (where we use the same weights as in the main analysis), the results are very similar– the only difference is that we drop observations that are within 250m of either treatment, giving us less power.

C.3 Impacts without randomization inference and inverse probability weights

In this subsection, we display the treatment effects we would have gotten if we were not to use inverse probability weights and randomization inference. Table C.6 displays these results. The direct treatment effects are generally smaller but the patterns are still similar. However, the spillover effects in these results are huge (.18SD for HSP, 0.3SD for MS). This shows that IPW's are crucial for getting the spillover effects right– the point estimates on the direct effects do not change as much because most segments have similar probabilities of being treated. However, there is a lot of heterogeneity in terms of likeliness of being a spillover unit as shown in section B.6.

Table C.4: Municipal services impacts on insecurity, pre-specified regressions

	ITT of assignment to:				
	Accounting for distant spillovers	Accounting for proximal spillovers			No spillovers
Dependent variable	MS outer spillover (1)	MS treated (2)	MS inner spillover (experimental) (3)	MS inner spillover (non-experimental) (4)	MS treated (5)
Insecurity index, z-score (+ more insecure)	-0.092	-0.118	0.094	-0.094	-0.156
	0.441	*0.147*	*0.196*	*0.447*	***0.024***
Perceived risk index, z-score (+ riskier)	-0.011	-0.119	0.042	-0.148	-0.133
	0.862	*0.123*	*0.516*	*0.243*	***0.037***
Crime index, z-score (+ more crime)	-0.141	-0.078	0.114	-0.009	-0.128
	0.136	*0.346*	*0.115*	*0.972*	***0.064***
Perceived & actual incidence of crime, z-score	-0.064	-0.108	0.067	0.011	-0.132
	0.612	*0.191*	*0.339*	*0.851*	***0.062***
# crimes reported to police on street segment	-0.254	0.002	0.177	-0.021	-0.088
	0.105	*0.956*	*0.130*	***0.023***	*0.493*

Notes: This table reports intent to treat (ITT) estimates of the effects of municipal services using the pre-specified regressions. Randomization inference p-values are italicized.

Table C.5: Interaction impacts on insecurity, pre-specified regressions

	ITT of assignment to:					
	Accounting for proximal spillovers			No spillovers		
Dependent variable	HSP effect (2)	MS effect (3)	Interaction effect (4)	HSP effect (5)	MS effect (6)	Interaction effect (7)
Insecurity index, z-score (+ more insecure)	-0.186	-0.169	0.061	-0.084	-0.079	-0.187
	0.089	*0.212*	*0.661*	*0.153*	*0.378*	*0.218*
Perceived risk index, z-score (+ riskier)	-0.194	-0.171	0.054	-0.094	-0.057	-0.145
	0.080	*0.184*	*0.687*	*0.112*	*0.520*	*0.357*
Crime index, z-score (+ more crime)	-0.117	-0.111	0.048	-0.045	-0.074	-0.167
	0.298	*0.427*	*0.725*	*0.489*	*0.407*	*0.291*
Perceived & actual incidence of crime, z-score	-0.162	-0.249	0.238	-0.061	-0.141	0.038
	0.199	*0.096*	*0.213*	*0.310*	*0.144*	*0.757*
# crimes reported to police on street segment	-0.007	0.181	-0.321	-0.005	0.073	-0.515
	0.915	*0.416*	*0.326*	*0.955*	*0.653*	*0.067*

Notes: This table reports intent to treat (ITT) estimates of the effects of both interventions using the pre-specified regressions. Randomization inference p-values are italicized.

Thus estimating unbiased treatment and spillover effects in the presence of the geographic clustering of high crime areas requires the use of inverse probability weights and randomization inference.

C.4 Program effects assuming spillovers and no interaction

Table C.7 displays program impacts on security in the experimental sample, accounting for proximal spillovers and no interaction between treatments.

C.5 Robustness to weight top-coding

We pre-specified top-coding our weights at 20 so that observations are not given undue weight. In table C.8, we conduct a sensitivity analysis with and without top-coding to test the robustness of our results to this decision. We display the results when we top-code the weights at six different values plus a version without any top-coding. Although the coefficients move around a bit, they point in the same direction as our main set of results. Therefore, the decision to top-code our weights is not driving our main set of results.

C.6 Station-level results

Table C.9 displays impacts on insecurity by station. Stations match *localidades* in Bogotá, which are administrative units for other planning purposes as the provision of municipal services. They also have independent budgets for some specific public services. Since we blocked by police station, technically each police station is an "almost" independent experiment. They are not completely independent as the interventions in one station may interfere with crime or treatment in a neighboring station.

The largest and more robust decreases in insecurity are in the Chapinero, Kennedy and Engativá police stations. Chapinero is a central police station known to have very heavy traffic and moving population. It also concentrates the financial center of the city and the country. Kennedy and Engativá are at the west side of the city and connect it to the west part of Colombia as well as the Caribbean coast. Eldorado International Airport is located between these two stations, and most of the traffic to Medellín and Barranquilla, the second and fourth largest cities in Colombia has to traverse this part of Bogotá.

On the other hand, there is a statistically significant increase in crime in the Mártires police station. This result may be explained because that is precisely where the Bronx is located, the only part of the city where the government did not have permanent presence and was subject of a large intervention when both the hot spots policing and the municipal services treatments were being delivered. This large intervention disrupted the presence of homeless drug users throughout the jurisdiction of the station.

C.7 Aggregate effects with on crime subgroups

In tables C.10, C.11 and C.12, we display the aggregate effects on crime subgroups with confidence intervals.

C.8 Aggregate effects with different radii

Table C.13 presents the estimation of aggregate effects using 125 meters radius. Generally, we see similar results as when using radius of 250 meters: a decrease in crime in targeted hot spots (in this case between 24 and 132 total deterred crimes, depending on the specification), and spillovers to both experimental and non-experimental streets. The aggregate effects is an increase of 50 to 60 crimes, although rather imprecise (note the confidence intervals are virtually centered at 0).

Table C.6: Naïve treatment effects

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both	Sum of (2), (3), and (4)	Impact of proximal spillovers: Any intensive policing	Any municipal services	Both	Sum of (6), (6), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Insecurity index, z-score (+ more insecure)	0.066	0.005	-0.071	-0.094	-0.160	0.177	0.301	-0.221	0.257
		[.057]	[.091]	[.130]	[.102]	[.064]***	[.066]***	[.096]**	[.079]***
Perceived risk index, z-score (+ riskier)	-0.018	-0.022	-0.080	-0.105	-0.207	0.104	0.205	-0.187	0.122
		[.058]	[.088]	[.131]	[.103]**	[.063]	[.064]***	[.089]**	[.076]
Crime index, z-score (+ more crime)	0.128	0.030	-0.037	-0.053	-0.060	0.191	0.296	-0.181	0.305
		[.055]	[.095]	[.135]	[.103]	[.067]***	[.067]***	[.105]*	[.084]***
Perceived & actual incidence of crime, z-score	0.002	-0.001	-0.119	0.095	-0.025	0.143	0.236	-0.208	0.171
		[.062]	[.090]	[.143]	[.119]	[.067]**	[.071]***	[.097]**	[.083]**
# crimes reported to police on street segment	1.334	0.081	0.128	-0.320	-0.112	0.233	0.337	-0.087	0.484
		[.091]	[.160]	[.218]	[.164]	[.124]*	[.115]***	[.196]	[.164]***

Notes: This table reports intent to treat (ITT) estimates of the effects of the two interventions, via a OLS regression of each outcome on treatment indicators, police station (block) fixed effects, and baseline covariates (see equation **??**. Column 5 reports the sum of the three treatment coefficients. Standard errors are clustered using the following rules: For all treated segments except with cluster size 2, each segment is a cluster. For all other untreated segments, each segment gets its own cluster ID. For entirely untreated quadrants, they form a cluster. For quadrants with exactly 2 units assigned to treatment, those units form a cluster.

Table C.7: Program impacts on security in the experimental sample, accounting for proximal spillovers (with p-values from randomization inference)

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	ITT of assignment to: Any municipal services	ITT of assignment to: Sum of (2) and (3)	Impact of proximal spillovers: Any intensive policing	Impact of proximal spillovers: Any municipal services	Impact of proximal spillovers: Sum of (5) and (6)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Insecurity index, z-score (+ more insecure)	0.003	-0.143	-0.110	-0.253	-0.026	0.111	0.085
		0.270	*0.215*	*0.107*	*0.703*	***0.064***	*0.218*
Perceived risk index, z-score (+ riskier)	0.049	-0.149	-0.090	-0.239	-0.048	0.056	0.008
		0.174	*0.289*	***0.091***	*0.906*	*0.226*	*0.469*
Crime index, z-score (+ more crime)	-0.044	-0.088	-0.093	-0.182	0.005	0.129	0.135
		0.537	*0.314*	*0.261*	*0.564*	***0.037***	*0.109*
Perceived & actual incidence of crime, z-score	-0.013	-0.071	-0.101	-0.173	-0.015	0.074	0.059
		0.683	*0.224*	*0.286*	*0.683*	*0.162*	*0.288*
# crimes reported to police on street segment	0.138	-0.100	-0.053	-0.153	0.043	0.204	0.246
		0.558	*0.824*	*0.555*	*0.604*	***0.053***	*0.100*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports intent to treat (ITT) estimates of equation 1, estimating the direct effects of the two interventions (Columns 2 to 4) and the spillover effects (Columns 5 to 7) via a WLS regression of each outcome on treatment indicators, spillover indicators, police station (block) fixed effects, and baseline covariates. Columns 4 and 7 report the sum of the two preceding coefficients. The measures of perceived risk, perceived incidence of crime, and proportion reporting crime come from our citizen survey, and the # of crimes reported to the police come from police administrative data.

Table C.8: Robustness to weight top-coding

	Outcome: Index of insecurity						
	No	Weights top-coded at:					
	top-coding	5	10	15	20	25	50
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: With interaction effect							
Assigned to hotspot policing (HSP)	-0.046	-0.066	-0.054	-0.049	-0.049	-0.047	-0.046
	[.055]	[.050]	[.053]	[.054]	[.055]	[.055]	[.055]
Assigned to municipal services (MS)	-0.066	-0.099	-0.084	-0.074	-0.07	-0.068	-0.066
	[.088]	[.088]	[.087]	[.088]	[.088]	[.088]	[.088]
Interaction effect (IE)	-0.242	-0.127	-0.150	-0.174	-0.199	-0.216	-0.242
	[.133]*	[.130]	[.129]	[.129]	[.130]	[.131]*	[.133]*
Effect of receiving both (HSP+MS+IE)	-0.354	-0.292	-0.288	-0.297	-0.318	-0.332	-0.354
	[.100]***	[.097]***	[.096]***	[.095]***	[.095]***	[.097]***	[.100]***
Panel B: Without interaction effect							
Assigned to hotspot policing (HSP)	-0.144	-0.092	-0.101	-0.113	-0.126	-0.133	-0.144
	[.062]**	[.049]*	[.054]*	[.058]*	[.060]**	[.061]**	[.062]**
Assigned to municipal services (MS)	-0.183	-0.150	-0.148	-0.153	-0.163	-0.171	-0.183
	[.069]***	[.067]**	[.066]**	[.067]**	[.067]**	[.068]**	[.069]***
Effect of receiving both (HSP+MS)	-0.327	-0.241	-0.249	-0.266	-0.289	-0.304	-0.327
	[.093]***	[.079]***	[.082]***	[.084]***	[.087]***	[.089]***	[.093]***

Notes: This figure displays treatment effects on the index of insecurity assuming no spillovers. In column 1, we do not top-code the weights. In columns 2–7, we top code at the value listed above. The tables in the main text are top-coded at 20, as pre-specified in our pre-analysis plan.

Table C.9: Impacts on insecurity, by station

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)	N
	(1)	(2)	(3)	(4)	(5)	(6)
Insecurity index, z-score	0.078	-0.049	-0.070	-0.199	-0.318	1916
(+ more insecure)		[.055]	[.088]	[.130]	[.095]***	
by station:						
Usaquen	-0.996	0.526	0.138	-0.660	0.004	113
		[.210]**	[.503]	[.609]	[.348]	
Chapinero	-0.256	-0.025	0.028	-0.932	-0.928	268
		[.187]	[.279]	[.417]**	[.239]***	
Santa Fe	0.222	0.480	-0.262	0.315	0.534	86
		[.214]**	[.297]	[.468]	[.377]	
San Cristobal	0.696	-0.306	-0.822	0.466	-0.662	48
		[.441]	[.492]	[.760]	[.585]	
Usme	0.838	0.172	-0.201	0.505	0.476	31
		[.416]	[.421]	[.598]	[.644]	
Tunjuelito	1.110	0.211	0.102	-0.076	0.237	49
		[.326]	[.410]	[.532]	[.503]	
Bosa	0.423	-0.394	-0.490	0.473	-0.411	58
		[.271]	[.572]	[.696]	[.406]	
Kennedy	0.274	-0.054	-0.191	-0.236	-0.482	197
		[.138]	[.262]	[.323]	[.194]**	
Fontibon	0.040	0.077	0.398	-1.128	-0.653	106
		[.236]	[.392]	[.643]*	[.539]	
Engativa	-0.043	-0.023	-0.111	-0.751	-0.885	100
		[.233]	[.220]	[.438]*	[.361]**	
Suba	-0.195	-0.076	0.183	-0.308	-0.200	237
		[.120]	[.260]	[.325]	[.204]	
Barrios Unidos	-0.448	-0.125	0.127	-0.078	-0.076	64
		[.139]	[.230]	[.247]	[.232]	
Teusaquillo	0.076	-0.154	0.573	-0.261	0.158	63
		[.370]	[.358]	[.524]	[.284]	
Martires	0.706	-0.295	-0.476	1.459	0.688	86
		[.226]	[.236]**	[.318]***	[.288]**	
Antonio Narino	0.661	-0.150	0.106	-0.463	-0.507	51
		[.316]	[.615]	[.968]	[.427]	
Puente Aranda	-0.290	0.485	-0.196	-0.447	-0.158	45
		[.616]	[.416]	[.651]	[.622]	
Candelaria	0.056	0.133	-0.090	-0.557	-0.514	60
		[.364]	[.519]	[.627]	[.523]	
Rafael Uribe	0.791	-0.192	-0.122	0.014	-0.299	101
		[.169]	[.177]	[.354]	[.301]	
Ciudad Bolivar	0.176	-0.182	-0.257	0.316	-0.123	153
		[.219]	[.289]	[.461]	[.377]	

Notes: This table reports intent to treat (ITT) estimates of the effects of the two interventions by station. The treatment effects report the marginal effect of receiving any treatment or of both, and Column 5 reports the sum of the three treatment coefficients. Standard errors are clustered using the following rules: (i) for all treated segments except with cluster size 2, each segment is a cluster; (ii) for all other untreated segments, each segment gets its own cluster identifier; (iii) for entirely untreated quadrants, they form a cluster; and (iv) for quadrants with exactly 2 units assigned to treatment, those units form a cluster.

Table C.10: Estimated aggregate violent crime impacts of the interventions, accounting for proximal spillovers in the experimental and non-experimental samples

	Dependent variable: # of violent crimes reported to police on segment (administrative data)							
	No interaction between treatments				Interaction between treatments			
Impacts of treatment	Coeff. (1)	RI p-value (2)	# segments (3)	Estimated total impact = (1) × (3) (4)	Coeff. (5)	RI p-value (6)	# segments (7)	Estimated total impact = (5) × (7) (8)
A. Direct treatment effect								
Intensive policing	-0.104	***0.036***	756	-78.3	-0.081	*0.101*	756	-61.4
Municipal services	-0.037	*0.539*	201	-7.3	0.000	*0.941*	201	0.0
Both					-0.097	*0.334*	75	-7.3
Subtotal				-85.6				-68.7
B. Spillover, experimental sample								
Intensive policing	-0.061	*0.221*	705	-43.0	-0.034	*0.573*	705	-24.2
Municipal services	0.011	*0.733*	546	5.8	0.037	*0.418*	546	20.2
Both					-0.088	*0.235*	281	-24.8
Subtotal				-37.3				-28.7
C. Spillover, non-experimental sample								
Intensive policing	0.003	***0.414***	51390	131.4	-0.004	*0.815*	51390	-184.3
Municipal services	-0.010	***0.087***	20740	-214.4	-0.017	***0.072***	20740	-358.7
Both					0.013	*0.237*	15491	199.5
Subtotal				-83.0				-343.5
Net increase (decrease) in crime				-205.9				-440.9
			95% CI	(-813, 394)			*95% CI*	(-1053, 147)
			90% CI	(-722, 303)			*90% CI*	(-963, 66)

Notes: This table presents the aggregate effect calculation for both interventions assuming proximal spillovers. Columns 1–4 refer to the non-interacted results (equation 1 under the constraint that $\beta_3 = 0$ and $\lambda_3 = 0$) while columns 5–8 refer to the interacted results (equation 1 with no constraints). Columns 1 and 5 display the bias-adjusted treatment effect while columns 2 and 6 display RI p-values. Columns 3 and 7 display the number of units in each group. Columns 4 and 8 display the product of the bias-adjusted treatment effect and the number of units in each group. The confidence interval on the bottom of the table is constructed using randomization inference. First we create a fake schedule of potential outcomes for each observation by adding or subtracting RI-adjusted treatment or spillover effects. This process gives us a potential outcome for each unit depending on its treatment assignment. Second, we simulate a randomization and take the potential outcome associated with the treatment assignment of the new randomization. Third we estimate treatment and spillover effects using this new outcome and apply the RI bias adjustment from our main set of results. Fourth, we multiply these bias-adjusted treatment effects by the number of segments in each group, and sum across both the experimental and non-experimental samples to get the aggregate effect. We repeat steps two through four 1,000 times to get the distribution of the test statistic, which is roughly centered on the actual number of deterred crimes. The 2.5 and 97.5 percentiles of this distribution give us the 95% confidence interval. $p < .1$ in bold.

Table C.11: Estimated aggregate property crime impacts of the interventions, accounting for proximal spillovers in the experimental and non-experimental samples

	Dependent variable: # of property crimes reported to police on segment (administrative data)							
	No interaction between treatments				Interaction between treatments			
	Coeff.	RI p-value	# segments	Estimated total impact = (1) × (3)	Coeff.	RI p-value	# segments	Estimated total impact = (5) × (7)
Impacts of treatment	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Direct treatment effect								
Intensive policing	0.003	*0.837*	756	2.5	0.085	*0.382*	756	64.1
Municipal services	-0.016	*0.998*	201	-3.3	0.111	*0.387*	201	22.4
Both					-0.339	*0.145*	75	-25.4
Subtotal				-0.8				61.0
B. Spillover, experimental sample								
Intensive policing	0.104	*0.314*	705	73.1	0.158	*0.187*	705	111.1
Municipal services	0.193	***0.039***	546	105.4	0.243	***0.040***	546	132.7
Both					-0.179	*0.375*	281	-50.4
Subtotal				178.6				193.4
C. Spillover, non-experimental sample								
Intensive policing	0.015	***0.091***	51390	776.4	0.021	***0.031***	51390	1,086.2
Municipal services	-0.007	*0.744*	20740	-142.5	0.000	*0.815*	20740	-1.0
Both					-0.013	*0.477*	15491	-197.0
Subtotal				633.8				888.2
Net increase (decrease) in crime				811.6				1,142.7
			95% CI	(-436, 1951)			*95% CI*	(-149, 2357)
			90% CI	(-272, 1802)			*90% CI*	(48, 2126)

Notes: This table presents the aggregate effect calculation for both interventions assuming proximal spillovers. Columns 1–4 refer to the non-interacted results (equation 1 under the constraint that $\beta_3 = 0$ and $\lambda_3 = 0$) while columns 5–8 refer to the interacted results (equation 1 with no constraints). Columns 1 and 5 display the bias-adjusted treatment effect while columns 2 and 6 display RI p-values. Columns 3 and 7 display the number of units in each group. Columns 4 and 8 display the product of the bias-adjusted treatment effect and the number of units in each group. The confidence interval on the bottom of the table is constructed using randomization inference. First we create a fake schedule of potential outcomes for each observation by adding or subtracting RI-adjusted treatment or spillover effects. This process gives us a potential outcome for each unit depending on its treatment assignment. Second, we simulate a randomization and take the potential outcome associated with the treatment assignment of the new randomization. Third we estimate treatment and spillover effects using this new outcome and apply the RI bias adjustment from our main set of results. Fourth, we multiply these bias-adjusted treatment effects by the number of segments in each group, and sum across both the experimental and non-experimental samples to get the aggregate effect. We repeat steps two through four 1,000 times to get the distribution of the test statistic, which is roughly centered on the actual number of deterred crimes. The 2.5 and 97.5 percentiles of this distribution give us the 95% confidence interval. $p < .1$ in bold.

Table C.12: Estimated aggregate homicide and rape impacts of the interventions, accounting for proximal spillovers in the experimental and non-experimental samples

	Dependent variable: # of homicides and rapes reported to police on segment (administrative data)							
	No interaction between treatments				Interaction between treatments			
	Coeff.	RI p-value	# segments	Estimated total impact = (1) × (3)	Coeff.	RI p-value	# segments	Estimated total impact = (5) × (7)
Impacts of treatment	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Direct treatment effect								
Intensive policing	-0.012	*0.254*	756	-9.0	-0.008	*0.418*	756	-6.1
Municipal services	-0.006	*0.683*	201	-1.3	0.000	*0.932*	201	0.0
Both					-0.017	*0.456*	75	-1.3
Subtotal				-10.2				-7.4
B. Spillover, experimental sample								
Intensive policing	0.005	*0.660*	705	3.3	0.011	*0.338*	705	7.9
Municipal services	-0.009	***0.382***	546	-5.0	-0.002	***0.872***	546	-1.4
Both					-0.022	*0.251*	281	-6.1
Subtotal				-1.8				0.4
C. Spillover, non-experimental sample								
Intensive policing	-0.001	***0.440***	51390	-59.5	-0.002	***0.144***	51390	-107.1
Municipal services	0.000	*0.849*	20740	-6.4	-0.001	*0.511*	20740	-27.8
Both					0.002	*0.379*	15491	31.8
Subtotal				-65.9				-103.1
Net increase (decrease) in crime				-78.0				-110.1
			95% CI	(-200,43)			*95% CI*	(-246,8)
			90% CI	(-177,21)			*90% CI*	(-220,-10)

Notes: This table presents the aggregate effect calculation for both interventions assuming proximal spillovers. Columns 1–4 refer to the non-interacted results (equation 1 under the constraint that $\beta_3 = 0$ and $\lambda_3 = 0$) while columns 5–8 refer to the interacted results (equation 1 with no constraints). Columns 1 and 5 display the bias-adjusted treatment effect while columns 2 and 6 display RI p-values. Columns 3 and 7 display the number of units in each group. Columns 4 and 8 display the product of the bias-adjusted treatment effect and the number of units in each group. The confidence interval on the bottom of the table is constructed using randomization inference. First we create a fake schedule of potential outcomes for each observation by adding or subtracting RI-adjusted treatment or spillover effects. This process gives us a potential outcome for each unit depending on its treatment assignment. Second, we simulate a randomization and take the potential outcome associated with the treatment assignment of the new randomization. Third we estimate treatment and spillover effects using this new outcome and apply the RI bias adjustment from our main set of results. Fourth, we multiply these bias-adjusted treatment effects by the number of segments in each group, and sum across both the experimental and non-experimental samples to get the aggregate effect. We repeat steps two through four 1,000 times to get the distribution of the test statistic, which is roughly centered on the actual number of deterred crimes. The 2.5 and 97.5 percentiles of this distribution give us the 95% confidence interval. $p < .1$ in bold.

Table C.13: Estimated aggregate impacts of the interventions, accounting for proximal spillovers in the experimental and non-experimental samples at 125 meters radius

	Dependent variable: # of crimes reported to police on segment (administrative data)							
	No interaction between treatments				Interaction between treatments			
Impacts of treatment	Coeff. (1)	RI p-value (2)	# segments (3)	Estimated total impact = (1) × (3) (4)	Coeff. (5)	RI p-value (6)	# segments (7)	Estimated total impact = (5) × (7) (8)
A. Direct treatment effect								
Intensive policing	-0.147	0.323	756	-111.3	-0.007	0.937	756	-5.6
Municipal services	-0.102	0.527	201	-20.6	0.088	0.531	201	17.6
Both					-0.485	**0.065**	75	-36.3
Subtotal				-131.9				-24.4
B. Spillover, experimental sample								
Intensive policing	-0.051	0.979	416	-21.2	-0.054	0.951	416	-22.4
Municipal services	0.162	0.179	251	40.8	0.144	0.314	251	36.3
Both					0.059	0.815	107	6.3
Subtotal				19.6				20.1
C. Spillover, non-experimental sample								
Intensive policing	0.012	0.208	22 693	270.1	0.007	0.282	22 693	162.9
Municipal services	-0.015	0.759	7 258	-108.6	-0.020	0.844	7 258	-146.0
Both					0.012	0.966	4 339	50.7
Subtotal				161.4				67.6
Net increase (decrease) in crime				49.1				63.4
			95% CI	(-749, 695)			95% CI	(-641, 680)
			90% CI	(-641, 559)			90% CI	(-572, 596)

Notes: This table presents the aggregate effect calculation for both interventions assuming proximal spillovers. Columns 1–4 refer to the non-interacted results (equation 1 under the constraint that $\beta_3 = 0$ and $\lambda_3 = 0$) while columns 5–8 refer to the interacted results (equation 1 with no constraints). Columns 1 and 5 display the bias-adjusted treatment effect while columns 2 and 6 display RI p-values. Columns 3 and 7 display the number of units in each group. Columns 4 and 8 display the product of the bias-adjusted treatment effect and the number of units in each group. The confidence interval on the bottom of the table is constructed using randomization inference. First we create a fake schedule of potential outcomes for each observation by adding or subtracting RI-adjusted treatment or spillover effects. This process gives us a potential outcome for each unit depending on its treatment assignment. Second, we simulate a randomization and take the potential outcome associated with the treatment assignment of the new randomization. Third we estimate treatment and spillover effects using this new outcome and apply the RI bias adjustment from our main set of results. Fourth, we multiply these bias-adjusted treatment effects by the number of segments in each group, and sum across both the experimental and non-experimental samples to get the aggregate effect. We repeat steps two through four 1,000 times to get the distribution of the test statistic, which is roughly centered on the actual number of deterred crimes. The 2.5 and 97.5 percentiles of this distribution give us the 95% confidence interval. $p < .1$ in bold.

C.9 Program impacts on police and mayor opinion assuming no interaction

Table C.14 displays program impacts on police and mayor opinion for the experimental sample when assuming no interaction term.

C.10 Treatment effects for 1 spillover case and different radii

In tables C.15 and C.16, we display program impacts on insecurity both the experimental and non-experimental samples, respectively. In both cases we account for spillovers within 125 meters and interaction effect between treatments. Generally we see the same patterns in the experimental sample, while for the non-experimental sample we see more robust evidence of short range spillovers within 125 meters than what we observe at 250 meters. This points in the direction of stronger displacement effects within shorter ranges.

In tables C.17 and C.18, we display program impacts on insecurity in the experimental and non-experimental samples, this time accounting for longer range spillovers within 500 meters. Generally, we see the same patters as we do for 250 meters, which suggests there is no change in crime levels between 250 and 500 meters.

C.11 Spillover effects using 2 spillover regions

Tables C.19 and C.20 present the estimation of spillover effects using equations 1 and 2 respectively. In each case, we include dummies for two spillover regions. The proximate spillover region is between 0 and 250 meters, and the distant spillover region is between 250 and 500 meters. In the experimental sample we see some evidence of distant crime spillovers of 0.16 and 0.12 standard deviations resulting from the policing and the municipal services interventions, each taken alone. These effects cancel out for streets that are exposed to spillovers from both treatments, as the marginal effect of the interaction is -0.24 standard deviations. Turning to the non-experimental sample, when we reach the largest sample size (panel a), we see no evidence of spillovers whatsoever. This holds true for the smaller sample of non-experimental street for which we have survey data.

Table C.14: Impacts on state legitimacy

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Sum of (2) and (3)
	(1)	(2)	(3)	(4)
Opinion of police, z-score (+ better)	-0.075	0.012	0.041	0.053
		[.060]	[.068]	[.093]
Level of police trust	1.154	-0.010	0.009	-0.001
		[.022]	[.026]	[.036]
Rating of work by police	1.306	0.020	0.018	0.039
		[.019]	[.021]	[.029]
Would aid police	1.636	-0.012	0.006	-0.006
		[.026]	[.031]	[.043]
Satisfied by police	1.114	0.012	0.011	0.023
		[.020]	[.023]	[.031]
Opinion of mayor, z-score (+ better)	-0.056	-0.125	-0.018	-0.143
		[.066]*	[.076]	[.099]
Level of mayor trust	0.924	-0.060	-0.029	-0.090
		[.025]**	[.029]	[.039]**
Rating of work by mayor	1.151	-0.042	0.004	-0.038
		[.021]**	[.024]	[.035]
Would aid mayor	1.413	-0.014	-0.006	-0.020
		[.028]	[.033]	[.044]
Satisfied by mayor	0.946	-0.025	0.014	-0.012
		[.021]	[.025]	[.034]

Notes: This table reports intent to treat (ITT) estimates of the effects of the two interventions, via a WLS regression of each outcome on treatment indicators, police station (block) fixed effects, and baseline covariates (see equation **??**. We ignore spillovers and interaction effects. Standard errors are clustered using the following rules: (i) for all treated segments except with cluster size 2, each segment is a cluster; (ii) for all other untreated segments, each segment gets its own cluster identifier; (iii) for entirely untreated quadrants, they form a cluster; and (iv) for quadrants with exactly 2 units assigned to treatment, those units form a cluster. All measures come from our citizen survey.

Table C.15: Program impacts in the experimental sample, accounting for spillovers within 125 meters (N=1,916)

		ITT of assignment to:				Impact of <125m spillovers:			
Dependent variable	Control mean	Any intensive policing	Any municipal services	Both	Sum of (2), (3), and (4)	Any intensive policing	Any municipal services	Both	Sum of (6), (7), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Insecurity index, z-score (+ more insecure)	0.089	-0.032	-0.013	-0.211	-0.255	-0.043	0.145	0.088	0.191
		0.782	*0.893*	*0.210*	***0.074***	*0.979*	***0.068***	*0.685*	***0.084***
Perceived risk index, z-score (+ riskier)	0.016	-0.057	0.006	-0.165	-0.216	-0.060	0.150	0.081	0.171
		0.490	*0.996*	*0.304*	***0.096***	*0.670*	***0.051***	*0.645*	*0.118*
Crime index, z-score (+ more crime)	0.133	0.004	-0.026	-0.186	-0.208	-0.011	0.092	0.066	0.147
		0.880	*0.795*	*0.268*	*0.163*	*0.690*	*0.198*	*0.826*	*0.145*
Perceived & actual incidence of crime, z-score	-0.004	0.010	-0.083	-0.004	-0.077	0.014	0.053	0.061	0.127
		0.873	*0.353*	*0.891*	*0.632*	*0.529*	*0.313*	*0.930*	*0.216*
# crimes reported to police on street segment	1.359	-0.007	0.088	-0.485	-0.405	-0.054	0.144	0.059	0.149
		0.937	*0.531*	***0.065***	***0.090***	*0.951*	*0.314*	*0.815*	*0.303*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports intent to treat (ITT) estimates of equation 1, estimating the direct effects of the two interventions (Columns 2 to 4) and the spillover effects (Columns 6 to 8) via a WLS regression of each outcome on treatment indicators, spillover indicators, police station (block) fixed effects, and baseline covariates. Columns 5 and 9 report the sum of the three preceding coefficients. The measures of perceived risk, perceived incidence of crime, and proportion reporting crime come from our citizen survey, and the # of crimes reported to the police come from police administrative data.

Table C.16: Program impacts in the non-experimental sample, accounting for spillovers within125m from treated hot spots (N=399)

Dependent variable	Control mean	Impact of proximal spillovers: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)
	(1)	(2)	(3)	(4)	(5)
Insecurity index, z-score (+ more insecure)	-0.170	0.173	-0.136	0.664	0.702
		0.112	*0.607*	***0.048***	***0.007***
Perceived risk, z-score (+ riskier)	0.012	0.052	-0.244	0.957	0.764
		0.579	*0.375*	***0.015***	***0.009***
Crime incidence, z-score (+ more crime)	-0.294	0.237	0.019	0.150	0.405
		0.026	*0.892*	*0.515*	***0.035***
Perceived incidence of crime, z-score	-0.040	0.269	-0.025	0.286	0.530
		0.095	*0.916*	*0.382*	***0.047***
# crimes reported to police on street segment	0.294	0.118	0.096	-0.145	0.069
		0.114	*0.482*	*0.551*	*0.540*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports spillover effects in the non-experimental sample from equation 2, a WLS regression of each outcome on spillover indicators, police station (block) fixed effects, and baseline covariates 1. In panel (b), Column 5 reports the sum of the three spillover coefficients.

Table C.17: Program impacts in the experimental sample, accounting for spillovers within 500 meters (N=1,916)

		ITT of assignment to:				Impact of <500m spillovers:			
Dependent variable	Control mean	Any intensive policing	Any municipal services	Both	Sum of (2), (3), and (4)	Any intensive policing	Any municipal services	Both	Sum of (6), (7), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Insecurity index, z-score (+ more insecure)	-0.229	-0.060	-0.136	-0.081	-0.277	0.128	0.091	-0.204	0.016
		0.863	*0.383*	*0.714*	*0.607*	*0.129*	*0.167*	***0.055***	*0.467*
Perceived risk index, z-score (+ riskier)	-0.085	-0.118	-0.114	-0.047	-0.279	0.034	0.094	-0.182	-0.054
		0.833	*0.458*	*0.868*	*0.561*	*0.430*	*0.149*	***0.063***	*0.698*
Crime index, z-score (+ more crime)	-0.297	0.018	-0.113	-0.089	-0.183	0.179	0.058	-0.157	0.080
		0.499	*0.432*	*0.653*	*0.642*	***0.042***	*0.315*	*0.153*	*0.249*
Perceived & actual incidence of crime, z-score	-0.141	-0.003	-0.164	0.092	-0.075	0.132	0.014	-0.086	0.061
		0.627	*0.241*	*0.530*	*0.845*	*0.129*	*0.573*	*0.417*	*0.319*
# crimes reported to police on street segment	0.478	0.054	0.013	-0.411	-0.344	0.225	0.126	-0.255	0.096
		0.665	*0.834*	*0.135*	*0.186*	*0.132*	*0.303*	*0.186*	*0.448*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports intent to treat (ITT) estimates of equation 1, estimating the direct effects of the two interventions (Columns 2 to 4) and the spillover effects (Columns 6 to 8) via a WLS regression of each outcome on treatment indicators, spillover indicators, police station (block) fixed effects, and baseline covariates. Columns 5 and 9 report the sum of the three preceding coefficients. The measures of perceived risk, perceived incidence of crime, and proportion reporting crime come from our citizen survey, and the # of crimes reported to the police come from police administrative data.

Table C.18: Program impacts in the non-experimental sample, accounting for spillovers within 500m from treated hot spots (N=399)

Dependent variable	Control mean	Impact of proximal spillovers: Any intensive policing	Any municipal services	Both inter-ventions	Sum of (1), (2), and (3)
	(1)	(2)	(3)	(4)	(5)
Insecurity index, z-score (+ more insecure)	-0.239	0.042	-0.131	-0.026	-0.115
		0.710	*0.700*	*0.868*	*0.699*
Perceived risk, z-score (+ riskier)	-0.043	-0.042	-0.102	-0.020	-0.164
		0.936	*0.770*	*0.928*	*0.523*
Crime incidence, z-score (+ more crime)	-0.355	0.112	-0.116	-0.023	-0.027
		0.468	*0.710*	*0.848*	*0.995*
Perceived incidence of crime, z-score	-0.124	0.164	-0.116	-0.023	0.026
		0.447	*0.807*	*0.878*	*0.858*
# crimes reported to police on street segment	0.291	-0.016	-0.089	-0.016	-0.121
		0.943	*0.673*	*0.924*	*0.541*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports spillover effects in the non-experimental sample from equation 2, a WLS regression of each outcome on spillover indicators, police station (block) fixed effects, and baseline covariates 1. In panel (b), Column 5 reports the sum of the three spillover coefficients.

Table C.19: Spillover effects in the experimental sample, accounting for spillovers within 250 and between 250 and 500 meters (N=1,916)

		Impact of proximal spillovers:				Impact of distant spillovers:			
Dependent variable	Control mean	Any intensive policing	Any municipal services	Both	Sum of (2), (3), and (4)	Any intensive policing	Any municipal services	Both	Sum of (6), (7), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Insecurity index, z-score (+ more insecure)	-0.242	-0.057	-0.183	-0.107	-0.347	0.097	0.088	-0.228	-0.043
		0.858	*0.596*	*0.552*	*0.752*	*0.184*	*0.168*	***0.024***	*0.624*
Perceived risk index, z-score (+ riskier)	-0.098	-0.106	-0.171	-0.058	-0.335	0.001	0.027	-0.138	-0.110
		0.889	*0.698*	*0.811*	*0.783*	*0.515*	*0.370*	*0.120*	*0.786*
Crime index, z-score (+ more crime)	-0.305	0.010	-0.133	-0.120	-0.243	0.162	0.119	-0.242	0.039
		0.545	*0.552*	*0.443*	*0.666*	***0.074***	***0.078***	***0.019***	*0.377*
Perceived & actual incidence of crime, z-score	-0.149	-0.010	-0.203	0.091	-0.122	0.120	0.069	-0.160	0.029
		0.622	*0.277*	*0.296*	*0.841*	*0.160*	*0.226*	*0.110*	*0.432*
# crimes reported to police on street segment	0.473	0.046	0.032	-0.491	-0.413	0.201	0.184	-0.337	0.048
		0.740	*0.635*	*0.018*	***0.077***	*0.207*	*0.105*	***0.081***	*0.539*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports spillover estimates of equation 1 including two different spillover regions: 0-250m and 250-500m. Proximal spillovers are reported in columns 2-5 and distant spillovers are reported in columns 6-9. We estimate the coefficients via a WLS regression of each outcome on spillover indicators, police station (block) fixed effects, and baseline covariates. The measures of perceived risk, perceived incidence of crime, and proportion reporting crime come from our citizen survey, and the # of crimes reported to the police come from police administrative data.

Table C.20: Spillover effects in the non-experimental sample, accounting for spillovers within 250 and between 250 and 500 meters

		Impact of proximal spillovers:				Impact of distant spillovers:			
Dependent variable	Control mean	Any intensive policing	Any municipal services	Both	Sum of (2), (3), and (4)	Any intensive policing	Any municipal services	Both	Sum of (6), (7), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
a. Full sample (N=77,848)									
# crimes reported to police on street segment	0.201	0.007	0.000	0.009	0.016	0.006	0.010	0.008	0.024
		0.360	*0.861*	*0.970*	*0.243*	*0.424*	*0.546*	*0.779*	*0.172*
b. Surveyed sample (N=399)									
Insecurity index, z-score (+ more insecure)	-0.339	0.039	-0.097	-0.223	-0.280	-0.028	-0.103	-0.021	-0.152
		0.742	*0.683*	*0.356*	*0.200*	*1.000*	*0.611*	*1.000*	*0.783*
Perceived risk index, z-score (+ riskier)	-0.153	-0.042	-0.122	-0.203	-0.367	-0.055	-0.010	-0.012	-0.076
		0.910	*0.536*	*0.459*	***0.095***	*1.000*	*0.973*	*1.000*	*0.985*
Crime index, z-score (+ more crime)	-0.412	0.108	-0.040	-0.168	-0.100	0.007	-0.161	-0.023	-0.177
		0.457	*0.978*	*0.386*	*0.668*	*1.000*	*0.297*	*1.000*	*0.558*
Perceived & actual incidence of crime, z-score	-0.206	0.138	-0.042	-0.138	-0.043	0.026	-0.120	-0.052	-0.146
		0.518	*0.924*	*0.557*	*0.827*	*1.000*	*0.485*	*1.000*	*0.701*
# crimes reported to police on street segment	0.296	0.024	-0.026	-0.183	-0.185	-0.030	-0.200	0.037	-0.193
		0.704	*0.853*	*0.315*	*0.520*	*1.000*	*0.243*	*1.000*	*0.445*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports spillover estimates of equation 2 including two different spillover regions: 0-250m and 250-500m. Proximal spillovers are reported in columns 2-5 and distant spillovers are reported in columns 6-9. We estimate the coefficients via a WLS regression of each outcome on spillover indicators, police station (block) fixed effects, and baseline covariates. The measures of perceived risk, perceived incidence of crime, and proportion reporting crime come from our citizen survey, and the # of crimes reported to the police come from police administrative data.