NBER WORKING PAPER SERIES

PLACE-BASED INTERVENTIONS AT SCALE:
THE DIRECT AND SPILLOVER EFFECTS OF POLICING AND CITY SERVICES ON CRIME

Christopher Blattman
Donald Green
Daniel Ortega
Santiago Tobón

Working Paper 23941
http://www.nber.org/papers/w23941

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2017, Revised September 2018

Previously circulated as "Pushing Crime Around the Corner? Estimating Experimental Impacts of Large-Scale Security Interventions." This research is thanks to the collaboration of the National Police of Colombia and the Mayor's Office of Bogotá, especially Bogotá's 2016–18 Secretary of Security, Daniel Mejía, who co-conceived the experiment. Innovations for Poverty Action in Colombia and the Center for the Study of Security and Drugs at Universidad de los Andes coordinated research activities. Survey data was collected by Sistemas Especializados de Información. For research assistance we thank Juan Carlos Angulo, Peter Deffebach, Marta Carnelli, Daniela Collazos, Eduardo Garcia, Sofia Jaramillo, Richard Peck, Patryk Perkowski, Oscar Pocasangre, María Rodríguez, and Pablo Villar. For comments we thank Thomas Abt, Roseanna Ander, Anthony Braga, Adriana Camacho, Aaron Chalfin, Steve Durlauf, Marcela Eslava, Claudio Ferraz, Larry Katz, David Lam, Leopoldo Fergusson, Nicolás Grau, Sara Heller, Daniel Mejía, Ben Olken, Jan Pierskalla, Tristan Reed, Jacob N. Shapiro, Rodrigo Soares, Juan F. Vargas, David Weisburd, Dean Yang, and many conference and seminar participants. Data and analysis were funded by the J-PAL Governance Initiative, 3ie, the Development Bank of Latin America (CAF), Fundación Probogotá, Organización Ardila Lülle via Universidad de los Andes CESED, COLCIENCIAS in the Government of Colombia, and the J. William Fulbright Program. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Place-Based Interventions at Scale: The Direct and Spillover Effects of Policing and City Services on Crime
Christopher Blattman, Donald Green, Daniel Ortega, and Santiago Tobón
NBER Working Paper No. 23941
October 2017, Revised September 2018
JEL No. C93,K42,O10

ABSTRACT

Cities target police patrols and public services to control crime. What are the direct and spillover effects of such targeted state services? In 2016 the city of Bogotá, Colombia, experimented on an unprecedented scale. They randomly assigned 1,919 streets to either 8 months of doubled police patrols, greater municipal services, both, or neither. We study how crime responds to intensifying normal state presence in moderate-to high-crime streets, and what this implies about criminal behavior. Scale also brings challenges. Spatial spillovers in dense networks introduce bias and complicate variance estimation through "fuzzy clustering." But a design-based approach and randomization inference produce valid hypothesis tests in such settings. We find that increasing state presence has modest direct impacts, even when focusing on the highest-crime "hot spots." More intense state presence deters more crime. But in most cases, however, crime appears to displace to neighboring streets. Property crimes seem most easily displaced, while violent crimes may not be. One interpretation is that crimes with a more sustained motive are more likely to displace than crimes of passion, which state presence may more permanently deter.

Christopher Blattman
Harris School of Public Policy
The University of Chicago
1155 E 60th St.
Chicago, IL 60637
and NBER
blattman@uchicago.edu

Donald Green
Department of Political Science
Columbia University
420 W. 118 Street
7th Floor IAB
New York, NY 10027
dpg2110@columbia.edu

Daniel Ortega
Banco de Desarrollo de America Latina (CAF) &
Instituto de Estudios Superiores
de Administracion
Centro de Politicas Publicas
Ave. IESA, Edif. IESA
San Bernardino, Caracas 1010
Venezuela
dortega@caf.com

Santiago Tobón
University of Chicago
Pearson Institute and Harris Public Policy
& Innovations for Poverty Action
Peace and Recovery Program
Calle 98 No. 22-64 Of 307
Bogotá, Colombia
tobon@uchicago.edu

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/1156

1 Introduction

This paper takes advantage of two city-wide interventions in Bogotá, Colombia, to show that large-scale crime experiments are possible and, while they come with special econometric challenges, solving them not only shapes our views of common crime policies, but also the theories of criminal behavior that underlie them.

Police and city services are two of the most common tools of crime control. When crime rises, cities often respond by intensifying patrols or improving lighting in the affected areas. For example, more than 90% of United States police agencies use some form of "hot spots policing" that concentrates police on the highest crime streets. Another common tactic is to reduce physical and social disorder. Such "place-based" interventions focus on the locales where crime occurs rather than on the people responsible for them.[1]

If crime is closely coupled to particular places, then state presence will at least disperse the crime, and possibly deter it altogether. This question of displacement versus deterrence is not only crucial to evaluate the costs and benefits of the policies, it also has implications for our understanding of criminal incentives and behavior.

The current consensus is that place-based policies reduce aggregate crime. For example, increasing the intensity or quality of policing on high-crime hot spots appears to reduce crime on those corners, streets, or neighborhoods, as does tackling social disorder.[2] Moreover, two systematic reviews of the US literature find more instances of positive spillovers to nearby streets than negative ones.[3] These studies argue that place-based policing not only deter crimes, but that the benefits also diffuse to nearby streets.

What would this imply for a theory of criminal behavior? In economic models, criminals weigh the returns from committing crimes against the risk of capture and expected sanctions (e.g. Becker, 1968; Ehrlich, 1973; Chalfin and McCrary, 2017b). Place-based interventions

[1] On hot spots policing see Weisburd and Telep (2016); Police Executive Research Forum, (2008). On disorder studies see Braga et al. (2015). On place-based theory and policy effectiveness see Weisburd et al. (2012) and Abt and Winship (2016).

[2] Chalfin and McCrary (2017b) review the evidence on increased policing and find that more police are usually associated with falling crime city-wide. Looking at targeted hot spots interventions, a systematic review of hot spots policing identified 19 eligible studies (including 9 experiments). Among 25 tests of the core hypothesis, 20 report improvements in crime (Braga et al., 2012). These evaluations are largely in the U.S. Exceptions include ongoing experimental evaluations in Medellin (Collazos et al., 2018) and Trinidad and Tobago (Sherman et al., 2014). For tackling social disorder, evaluations of municipal services are relatively rare. (Braga et al., 2015) review interventions designed to tackle social and physical disorder, but the majority tend to be a policing strategy rather than attempts at urban renewal. There is some evidence that street lighting reduces crime (Farrington and Welsh, 2008). Cassidy et al. (2014) review five studies suggesting there is weak evidence that urban renewal reduces youth violence.

[3] See Braga et al. (2012); Weisburd and Telep (2016). Two natural experiments that put round-the-clock police in areas of London and Buenos Aires also found no evidence of spatial displacement (Draca et al., 2011; Di Tella and Schargrodsky, 2004).

can increase a criminal's perceived risk of detection and capture, and so they should be less likely to commit these purposeful, "motivated" crimes in that place.[4] Therefore if targeted, place-based policies reduce the *total* number of motivated and material crimes in a city, it suggests at least one of the following: that criminal rents are highly concentrated and unequally distributed within cities; that some offenders are resistant to moving crime locations; or that the supply of crime is elastic to the actual or perceived risk of apprehension in a small number of areas.[5]

Of course, many crimes do not have a sustained motive. Examples include drunken brawls outside a bar, or a sudden opportunity to mug someone. In these cases, any intervention that stops the crime from happening at that time or place could prevent it altogether. Therefore, if place-based interventions do indeed decrease total crime, then it might also point to the non-motivated, non-economic roots of many offenses.

Based on the current consensus, more countries are adopting place-based tactics. In Latin America, arguably the most violent region in the world, governments are especially eager adopters.[6] Colombia's two largest cities, Bogotá and Medellín, have put place-based tactics at the center of their security strategies in recent years.

We see a few reasons for caution, and argue that larger-scale studies are needed to assess the aggregate effects of place-based policies. First, the studied interventions vary widely in their nature and intensity. They range from round-the-clock police presence, to drug house invasions, to changes in the police-community relationship. Second, consistent with the theory outlined above, and a large literature in criminology, the same intervention may have different spillover effects on different kinds of crime (Weisburd et al., 2006). Third, studies vary in when and how they measure spillovers. Altogether, it may not be possible to draw general conclusions about crime spillovers just yet. Indeed, different interventions, outcomes, timing, and spillover measures could explain our fourth point: that the evidence on the direction of spillovers is mixed, with studies pointing both ways.[7]

[4]Police presence disrupts this crime or raises the risk of capture, while city services light dark areas or increase the number of people on the street. State presence may also signal order, telling criminals to stay away and citizens that the state is present—a version of the famous "broken windows" hypothesis (Wilson and Kelling, 1982; Apel, 2013). Note that "Broken windows policing" is sometimes used to describe intensive, zero tolerance policing. But more visible state presence and physical order should send similar signals.

[5]For articulations of these channels, see Clarke and Weisburd (1994); Weisburd et al. (2006); Chiba and Leong (2014).

[6]Latin America has 42 of the 50 most dangerous cities and a third of the world's homicides (see Consejo Ciudadano para la Seguridad Pública y Justicia Penal and Global Study on Homicide 2013). Major cities also have fewer police per person than the U.S. or Europe. Policymakers are interested in the returns to higher quality or quantity of policing. Muggah et al. (2016) document the adoption of hot spots policing tactics in different Latin American countries. Also see Abt and Winship (2016) for recommendations to U.S. international development agencies.

[7]One very wide review of 102 place-based interventions found indications of positive spillovers in a quarter

Finally, because of small samples and aggregation challenges, this question of displacement or diffusion is unsettled. The median hot spots policing study has fewer than 30 hot spots per treatment arm. Thus no one study can rule out large adverse spillovers.[8] Collectively, of course, existing studies have greater power, and the balance of evidence points to spillovers being beneficial. Even so, as we discuss below, there are good reasons to think the uncertainty around this average includes large negative and positive spillovers.[9] Thus above all, the literature needs more statistical power in more cases to speak to the question of crime displacement.

With these theoretical and empirical questions in mind we worked with the city of Bogotá to design a large-scale, multi-arm security experiment. Bogotá is a large, thriving, relatively rich and developed city in a middle-income country, now at peace. It has a professionalized and well-regarded police force. Thus lessons from Bogotá are potentially relevant to a range of US and global cities. Our intention was to take advantage of the unusual size of the experiment to identify more subtle spillovers over flexible ranges, and to test whether the interventions have different effects on different types of crime.

Scale and spillovers also bring econometric challenges to program evaluation. In particular, we illustrate how spillovers in dense networks bias treatment effects and complicate variance estimation through "fuzzy clustering." Thus we also designed this experiment to illustrate how two tools—study design and randomization inference—can be used to estimate spillovers flexibly, and produce valid hypothesis tests when standard methods do not.

Specifically, in January 2016 a new city government in Bogotá decided to experiment in nearly 2,000 moderate to high-crime streets by intensifying normal police patrolling and improving municipal services such as lighting and clean-up. This is the first attempt by a city to experiment at such a scale, especially with more than one intervention at once. Another advantage is that Bogotá has street-level, geo-referenced crime data on all 136,984 streets in the city, plus the location of every police patrol every 30 seconds for the year.

Interestingly, we draw similar conclusions whether we look at the full sample of experimental streets, or restrict our attention to the few hundred highest-crime hot spots. We find that intensifying patrols and municipal services slightly reduced property and violent crime on the targeted streets. We also find that crime, especially property crime, seems to displace

---

and adverse spillovers in a quarter Guerette and Bowers (2009). If we limit our analysis to higher-quality hot spots policing studies, of the 9 experimental and non-experimental studies with more than 15 control units, evidence of positive spillovers are more common: 3 report evidence of adverse spillovers and 6 report evidence of diffusion of benefits.

[8] At this scale, individual experiments are not powered to detect spillovers of 0.4 or 0.5 standard deviations in size (see Appendix A).

[9] See Section 8.1. This section explains why current meta-analysis confidence intervals might dramatically understate the uncertainty around aggregate spillovers.

nearby. We see no evidence of violent crimes displacing, however.[10] More state presence, or more forms of state presence, have the largest direct effects on crime reduction, especially in higher-crime places. But in aggregate the direct effect on crime is very small, and seems to be more than outweighed by property crime that spills over nearby.

This pattern is consistent with the idea that offenders with sustained motives (like theft) respond strategically to targeted state presence and relocate. Crimes of passion, however, may be more easily deterred.[11] This suggests the decision to use place-based tactics depends on local crime patterns and what policymakers think are the key crimes to deter.

The remainder of this introduction summarizes the design and conclusions in more detail. The experiment itself proceeded in two steps. First, we worked with the police to identify an experimental sample of 1,919 street segments with security concerns. A segment is a length of street between two intersections, a common unit of police attention (Weisburd et al., 2012). The city nearly doubled police patrol time on 756 of the 1,919 segments, giving an additional 77 minutes of time to streets that otherwise received 92 minutes of patrol time per day. Second, after intensive policing began, the city targeted 201 segments for municipal services. We randomized assignment to intensive policing, more municipal services, both, or neither. At this scale, we were ex-ante powered to detect direct effects of 0.15 standard deviations. We also monitor impacts onto the 77,848 segments within 250m, and are powered to detect spillovers as small as 0.02 standard deviations (with 80% power, see Appendix A).

Both interventions reallocated existing city resources. No new police or city contractors were added. Rather, within their patrol area (a quadrant), officers were told to double their time on two streets, in multiple visits. This intensive policing lasted eight months. With 130 segments per quadrant, there is little impact on patrolling on other segments—something we can confirm with geo-referenced data on patrols. These patrols went about their normal duties, interacting with citizens, and stopping and frisking suspicious people. Two months later, the city added a second experimental intervention over the first. To tackle social disorder, they sent city contractors to repair lights and clean up trash.

As a result, Bogotá's interventions are different from hot spots programs that intensely target drug houses or change the local approach to policing and community engagement. Rather, we estimate the relationship between security and the intensity of normal state presence. We do so on streets ranging from moderate to very high crime. The closest comparisons are a handful of US policing interventions that increased police time by roughly

[10]Broadly, we refer to property crimes for economically motivated property crimes. In this sense, a violent robbery is coded here as a property rather than a violent crime event.

[11]In a recent randomized trial in Bogotá, Nussio and Norza Cespedes (2018) find that information campaigns on the number of arrests at a specific location (i.e. objective information on the probability of apprehension) decrease reports on motivated crimes but not on crimes of passion (at treated places).

1-3 hours per day on 20–60 high-crime hot spots (Sherman and Weisburd, 1995; Telep et al., 2014; Taylor et al., 2011). To the best of our knowledge, however, Bogotá is the first attempt to evaluate normal service delivery on a wide range of streets, and the first to estimate aggregate effects on crime city-wide.[12]

As we indicate above, one challenge with experimenting at this scale is that spillovers interfere with clean causal identification. Specifically, treating one street can affect the outcomes of nearby control streets (for instance, if criminals move to nearby high-profit segments). We used a design-based approach to account for spatial spillovers flexibly, largely because we did not want to make strong assumptions about the structure of spillovers. Following our pre-analysis plan, we partitioned control streets by distance from treated streets: 0–250 meters (m), 250–500m, and >500m. By comparing outcomes across treatment and control categories, we can first test for local spillovers in the 0–250m and 250–500m regions, and then use unaffected regions as a control group for estimating direct treatment effects. We estimate spillovers into the non-experimental sample the same way. Moreover, we test whether we draw similar conclusions if we model spillovers with more structure, such as with continuous decay functions.

Because crime is not distributed evenly across the city, however, spillovers present further estimation challenges. By simulating the experiment many times, we show that the close proximity of experimental streets leads to hard-to-model patterns of "fuzzy clustering" (Abadie et al., 2016). In most randomizations, segments close to experimental streets tend to be assigned to the same spillover status. This biases estimated treatment effects and understates standard errors. Without a fixed geographic unit of clustering, we cannot use standard correction procedures. This is a common but under-explored problem with experiments in social or spatial networks. Whether we model spillovers flexibly or with decay functions, we use randomization inference to estimate exact p-values in these settings.

Our main outcome is police data on reported crimes, available for every street. But reported crimes are incomplete and reporting could be correlated with treatment. Thus we also conducted a survey of about 24,000 citizens, providing measures of unreported crimes, security perceptions, and attitudes to the state. These data also suggest that treatment does not affect the likelihood a crime is reported, which allow us to trust the crime data more.

First, we see only slight evidence that intensifying state presence improved security. Look-

[12] The closest comes from Essex, UK, where police increased their patrolling by a much smaller margin—roughly 10 minutes per day—in the 200 meter areas around the site of the prior week's burglaries. Blanes I Vidal and Mastrobuoni (2017) compare crime rates in these treated areas to the areas around previous burglaries and find no apparent effect on crime. Unlike the US studies, Essex is not a hot spots intervention, since there is no evidence that burglaries were persistent. But the Essex results are consistent with our own, where we fail to see large and statistically significant direct effects on crime except in the most intensively treated hot spots.

ing at each intervention individually, crime fell an average of 0.1 standard deviations from both intensive policing and municipal services. Accounting for spillovers, however, neither decrease is statistically significant. Nonetheless, when a street received both interventions reported crimes fell by 57%, statistically significant at the 1% level. It is possible that there are increasing returns to both police patrols and municipal services.

Importantly, if we restrict our attention to the highest-crime streets, we can simulate more targeted "hot spots" interventions. The same patterns hold. In fact, more intensive policing alone is not associated with decreases in crime in these highest-crime streets.

In aggregate, the direct effects of both treatments are modest. To understand true city-wide impacts, we would need a randomized trial across cities. However, we can approximate general equilibrium impacts by aggregating total direct effects and spatial spillovers over the eight months. If we consider crimes reported to the police alone, for instance, our highest estimate is that roughly 100 crimes were prevented in treatment segments with both interventions. Our main estimate, however, is that not even one crime was prevented in treatment segments over eight months, city-wide. Obviously none of these aggregate decreases are statistically significant.

Meanwhile, we see evidence of adverse spillovers. On a street-by-street basis, these are small in magnitude. With tens of thousands of nearby streets, however, small effects add up. Over the eight months of the intervention, our best estimate is that treatment increased the total number of crimes reported in the city by about 1,000, or 2.6%. A 90% confidence interval includes zero, and so this should not be taken as strong evidence of adverse effects. Even so, our results rule out more than a 1–2% improvement in city-wide crime.

Importantly, in our main specification it is property crime, as opposed to violent crime, that is displaced. If we add our estimates of crimes directly deterred on treated streets to crimes displaced, reported property crimes rise by 1,384 over the eight months, significant at the 90% level. By the same method, violent crimes fall 369, including a 7% decline in homicides and sexual assaults (86 over eight months), though not statistically significant.

Displacement of property crime and deterrence of violent crime is consistent with standard economic models of crime: increasing the risk of detection stops criminals from committing motivated crimes in that specific place, but most likely the crime is not deterred and rather committed elsewhere. But crimes of passion, once avoided, may be less likely to sustain their motive and be displaced.

As we discussed, the evidence from US cities has tended not to find many adverse spillovers, at least within a 1–2 block radius. But several recent, large-scale, non-US studies tell similar stories to Bogotá. For example, a large-scale trial of intensive policing in another Colombian city, Medellin, draws similar conclusions—small direct effects and no evidence of beneficial spillovers, with wide confidence intervals for aggregate effects including the possibility of adverse spillovers (Collazos et al., 2018). In Mexico, Dell (2015) finds that drug trafficking, a crime with extremely strong and sustained motives, displaces to nearby municipalities in response to increased enforcement. Drunk driving is another criminal behavior that, once underway, may be impossible to deter. A recent large-scale experiment of drunk driving checkpoints in India shows displacement as drunk drivers take alternate routes (Banerjee et al., 2017).

Finally, this study offers a chance to demonstrate advances in accounting for spatial spillovers. First, economists have tended to impose a fair degree of structure on spillovers. Where the nature of spillovers is unknown, however, a more flexible design-based approach is more appropriate (Gerber and Green, 2012; Aronow and Samii, 2013; Vazquez-Bare, 2017). Second, standard methods overstate precision when the spillovers lead to fuzzy clusters. Randomization inference, seldom used in economics, provides valid hypothesis tests.

As more interventions go to scale in close proximity, these econometric approaches to place-based program evaluation and hard-to-model spillovers will only grow in importance. These problems and solutions are applicable to a variety of issues beyond crime. Many urban programs are both place-based and vulnerable to spillovers. This includes efforts to improve traffic flow, beautify blighted streets and properties, foster community mobilization, and rezone land use. The same challenges could arise with experiments in social and family networks (Abadie et al., 2016; Vazquez-Bare, 2017). Experiments in dense interrelated networks present a textbook case of where design-based and randomization inference needs to enter the econometric program evaluation toolkit.

2 Setting

Crime is one of the most pressing social problems in Bogotá, a middle income city of roughly 8 million people.[13] In the 1990s Bogotá was one of the most violent cities in the world, with 81 murders per 100,000 people. By 2016 the figure had fallen to 15.6, comparable to a large US city such as Chicago.[14]

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

[13] Bogotá had a 2015 GDP per capita of roughly $22,000 in purchasing power parity (PPP) terms. 10% of the population was below the national poverty line for metropolitan areas of PPP$6 a day.

[14] This is much lower than the most violent cities in the world, such as 120 in Caracas, 65 in Cape Town, 64 in Detroit, and 64 in Cali, Colombia. It is comparable in crime rates to a U.S. city like Chicago, with 15 murders per 100,000 in 2015, but greater than the 7 recorded in Los Angeles or 4 in New York. U.S. figures come from the FBI Uniform Crime Report and others from the World Atlas.

violence. Most violent crimes appear to be crimes of passion. The Mayor's office estimates that 81% of all the homicides in the city in 2015 were a result of fights, 12% were contract killings, and 5% were violent robberies. Finally, most offenders are individual young people. There are some semi-organized youth gangs, and some organized crime, but they do not seem to be responsible for the vast majority of the street crime or violence.

Like many cities, crime in Bogotá is also fairly 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. These higher-crime streets are distributed around the city. They include wealthy areas where criminals come to mug pedestrians, burgle homes, or steal expensive cars, as well as more barren industrial areas with little traffic, where it is easier to sell drugs or steal. They also include popular nightlife areas.

Security policy and policing Bogotá has moderate to low levels of police compared to large US and Latin American cities. Bogotá has about 18,000 police officers in operational activities, including about 6,200 patrol agents. We estimate about 239 police per 10,000 people. The Colombian average is 350, and most cities are above Bogotá's ratio. The national US ratio was 230 in 2013 but it is greater in large cities, including 413 in New York, 444 in Chicago, 611 in Washington, or 257 in Los Angeles.[15]

We discreetly observed police patrols and qualitatively interviewed residents on 100 of the treated streets, as described below. Our assessment is that patrols are reasonably well-regarded. The broader police force is not without problems, but street patrol officers are generally regarded as competent and non-corrupt. If anything, residents complained that officers were not present often enough. A survey of 24,000 residents, also discussed below, confirms these impressions of police patrols.

In January 2016 a new mayor came to power, Enrique Peñalosa. Crime reduction and increasing trust in government were central to his platform. In his first 100 days, the Mayor pledged to dedicate more municipal services and law enforcement in 750 high-crime street segments.

Municipal services included trash collection, tree pruning, graffiti clean-up, and streetlight maintenance. The performing agencies report directly to the Mayor's office, but the Mayor's power is limited by contracts and difficulties in monitoring and enforcing instructions.

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

[15]Data for Colombia was reported by the Secretariat of Security 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.

branch of the National Police and report up to the Minister of Defense. But the city has the power of the purse, as it pays for police equipment. The Colombian Constitution also calls on police to comply with the Mayors' requests and policies. Changes in force levels are much more expensive, however, and the national government rejected the Mayor's request to increase the number of police. Thus the Mayor's office focused on increasing police efficiency and quality, especially street patrols.

Police patrolling The quadrant (*cuadrante*) is the basic patrolling unit. Bogotá has 19 urban police stations. Stations are divided into CAIs—*Comando de Atención Inmediata*—a small local police base that coordinates patrol agents and takes civilian calls. Each CAI has about 10 quadrants. There are 1,051 quadrants, each with 130 street segments on average.

Each quadrant has six permanent patrol officers. They patrol in pairs, on motorbike and foot, in three shifts of eight hours each. In practice, patrols are expected to move about throughout their shift, by motorbike. They may patrol a street on motorbike or dismount 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. Patrols have daily quotas. They are expected to regularly stop and frisk any suspicious people, and will seize illegal weapons (usually knives) and other contraband. Patrols tend to focus interrogations on young men. An arrest means both patrollers must take the suspect to the station, for hours of paperwork and processing. This keeps them from meeting performance goals, and so patrols may avoid minor arrests.

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.

3 Interventions

In January 2016, a new city government came to power. A key plank of the Mayor's election platform was to identify the highest crime streets in the city and target them with a greater share of normal city services, especially police patrols and municipal services. We can view both interventions as an intensification of normal service delivery.

No new funds or personnel were added, and so this is effectively a randomized reallocation of city services. We are not concerned that control streets received materially fewer services as a result of the experiment. Treated streets are roughly 1% of all city streets, so increased attention to treated streets has only a tiny effect on control and non-experimental streets, on average.

Intensive policing Prior to the intervention, from 2012–15, normal police patrols spent roughly 10% of their time on the 2% highest-crime streets. Thus higher-crime streets already received a disproportionate amount of police attention. Nonetheless, these same streets recorded a quarter of all reported crimes in the city. Thus the intervention aimed to increase police time even further, in proportion to the crime they represented.

This intensive policing began on February 9, 2016 and ended on October 14, 2016.[16] It generally meant almost doubling police patrol time. As we will see below, during the intervention control streets received roughly 92 minutes of patrol time on average, with treated streets receiving an additional 77 minutes—an 84% increase.[17]

In order not to overextend patrols, the police required us to assign no more than two segments to treatment per quadrant so as not to distort regular duties too much. A 77-minute increase on two segments implied that patrol time fell on other segments in the quadrant by roughly one minute each. Thus we do not expect the reallocation to be a significant source of differential crime in treated, control, and nonexperimental streets.

Commanders told patrols to visit treatment segments at least 6 times per day for roughly 15 minutes each, mostly during the day unless near a bar. The police generally did not know what segments were in the control group, but in principle they could make reliable guesses. Commanders instructed patrols to continue their normal duties in treated segments: running criminal record checks; stopping, questioning, and frisking suspicious people; door-to-door visits to the community; conducting arrests or drug seizures; and so forth.[18]

Municipal services One city office coordinates street light maintenance and a second office is in charge of all clean-up activities. Both offices contract private companies to service

[16] The government, however, did not publicize the eligible high-crime streets, the existence of an experimental design, or which specific streets were being targeted. The Mayor's office initially planned to run this intensive policing intervention for at least 4 to 6 months. They extended the intervention in part to permit the research team enough time to fund and conduct a survey of citizens.

[17] Before the intervention, 1–2 weeks of GPS data suggested that experimental sample of streets received at least 38 minutes of patrol time per day. It is doubtful that actual time rose from 38 to 92 minutes. Rather, the 38 minutes is probably an understatement of average patrolling time per street, as there were fewer patrols with GPS devices patrolling city streets. The police did not have data on pre-intervention patrol times, since the GPS devices were piloted November 2015 through January 2016. See Appendix B.1.

[18] The only exception was in 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. Police cleared the streets and the city demolished the buildings. In this extreme case, it is obvious that more policing can reduce crime.

the streets. Contractors were expected to perform their usual duties, but the Mayor's office gave contractors lists of segments where they were asked to assess issues and deliver the appropriate services. The municipal services intervention began April 11, 2016 and continued until the end of the intensive policing intervention.

How do the Bogotá interventions compare to other interventions? Many of the US studies examine a change in policing approach rather than simply a change in intensity. These changes in approach vary widely. Some interventions take a "zero tolerance" approach, enforcing the most minor infractions. Others focus on "problem-oriented policing," where officers try to proactively address problems identified jointly with communities. Others place license plate readers on street corners, or crack down on drug corners and houses. The "hot spots" literature is, in short, a mixed bag of interventions that may or may not be directly comparable.

The Bogotá intervention is similar in style and approach to two US interventions that intensify patrol time but maintain normal duties, such as a Minneapolis study that raised patrol time to 3 hours per day on 55 hot spots (Sherman and Weisburd, 1995), a Jacksonville study where officers surveilled 78 hot spots for an additional 1–2 hours per day (Taylor et al., 2011). The rotation of 15-minute police patrols mirrors an intervention in Sacramento (Telep et al., 2014). Another example is an unpublished Medellín hot spots policing program, where 384 crime hot spots were treated with between 50 and 70 minutes of additional daily police patrolling time during six months in 2015 (Collazos et al., 2018).[19]

4 Experimental sample and design

4.1 Selecting the experimental sample

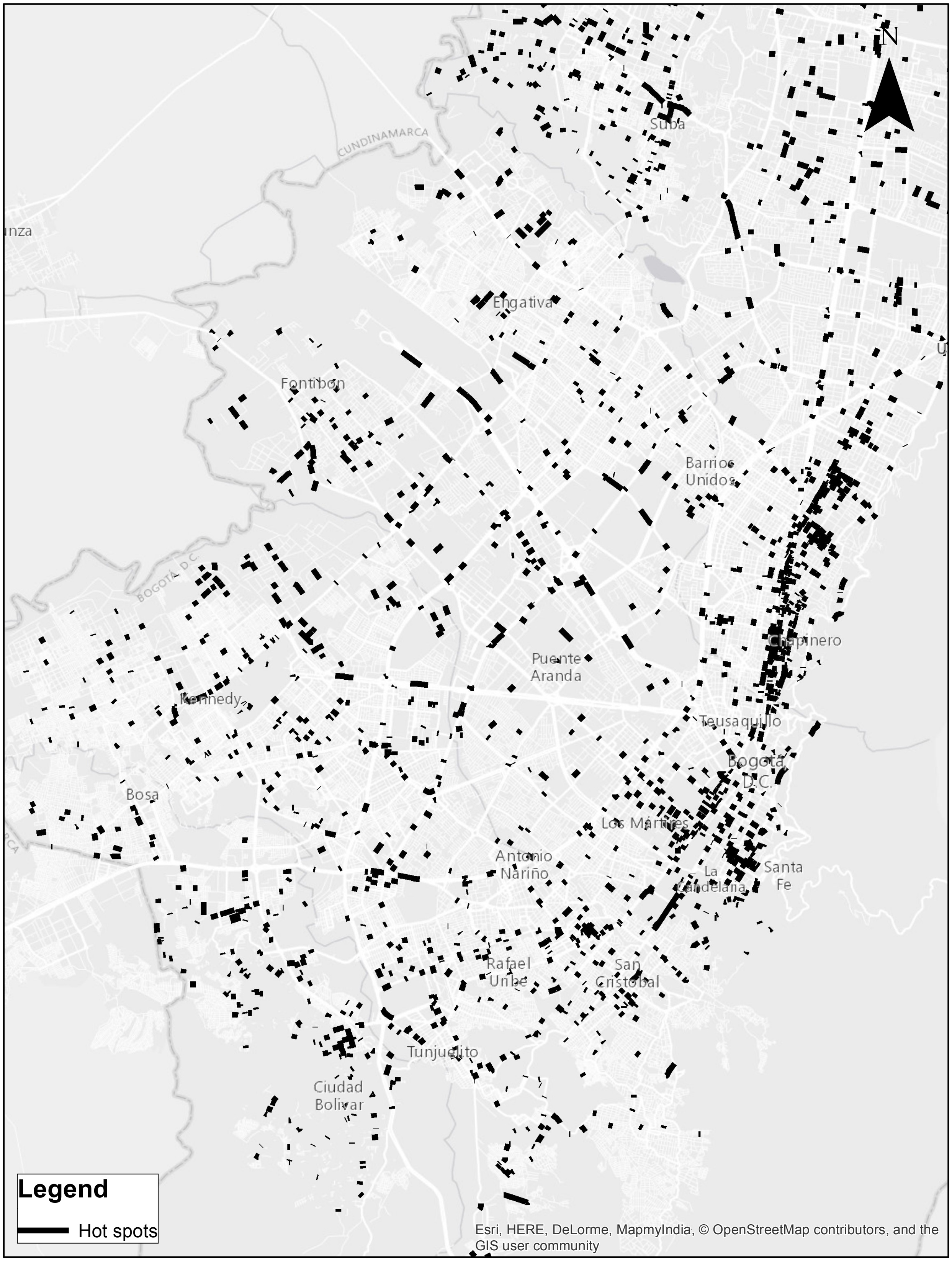
Figure 1 maps Bogotá's 136,984 street segments and indicates the 1,919 segments in our experimental sample.

To create this sample, the city started with the 2% highest-crime segments, using an index of reported crimes from geo-coded official statistics, between January 2012 and September 2015.[20] The city then asked each station's commanders and staff to verify the high-crime

[19]This Medellín study does not observe direct treatment effects on both property and violent crimes, although they do find evidence of a decrease in a particular form of crime: car thefts. They also find a decrease in car thefts in places nearby targeted segments. The context has some differences as well. For instance, while Medellín has about 60% more police per capita than Bogotá, the city also has highly organized criminal gang structures throughout the city, and police in these low and middle income neighborhoods may not be effective in deterring gang-associated crime because of the local power and influence of these groups.

[20]We constructed a geo-fence of 40m around each segment and assigned a reported crime to that segment

Figure 1: Map of experimental sample



Notes: Experimental street segments, in black, are the 1,919 streets included in our experimental sample.

segments. They did this partly because the geo-located crimes data were thought to contain errors (such as crimes assigned to the wrong street). Also, the official crime data are incomplete, omitting most unreported petty crimes and disorder. Calls or informal reports to police do not show up in official statistics, for instance, and police do not record crimes they observe or which people report informally.

Based on their knowledge, the police eliminated about a third of these segments, adding others in their stead, leaving 1,919 segments that account for 21% of the city's reported crimes.[21] As we discuss below, this led to an experimental sample with varying levels of crime, from low to acute. We will account for this in our analysis by looking at treatment effects in the highest reported crime streets, and also by attempting to measure smaller crimes and security perceptions with a citizen survey.

4.2 Design-based approach

We did not know the range of spatial spillovers, and so we pre-specified a flexible design that tested for spillovers in radii of 250m and 500m around treated streets.[22]

Failing to account for spillovers properly will bias treatment effect estimates. If control segments are close enough to treated segments 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 segments, and focused instead on the spillovers into nearby non-experimental segments. This is reasonable in small samples where hot spots are widely dispersed and the spillover regions do not overlap. But interference between units grows large as we scale up to hundreds of treated hot spots in a city. The same would be true of any intervention in a spatial or social network. This is a growing source of experimental work.

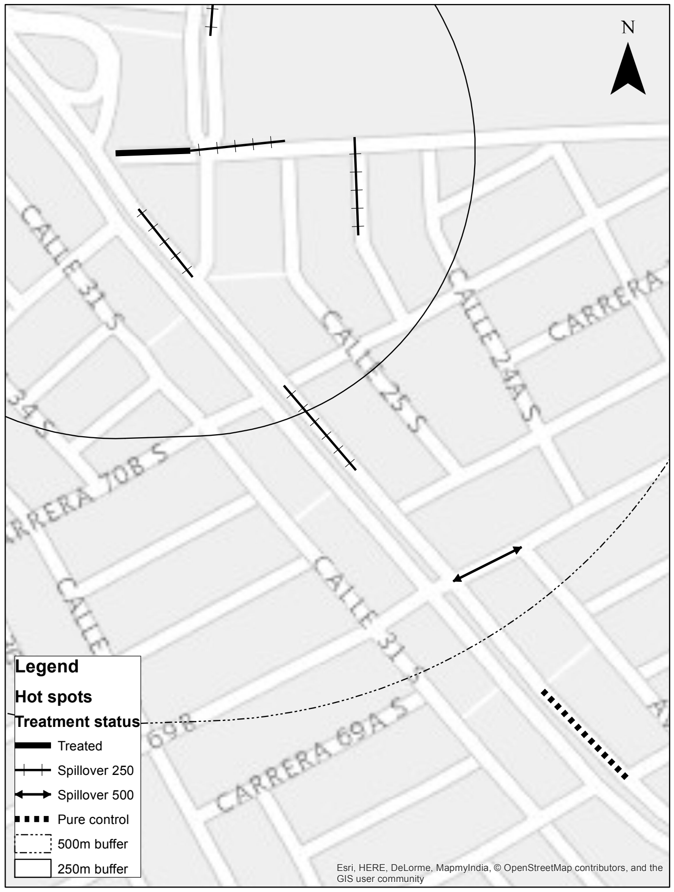
There are many ways to model spillovers. In economics it is common to use a continuous rate of decay. We will show the results of different continuous functions, but we felt that this imposed too much structure on the nature of spillovers. After all, crime might more easily displace to an opportune segment a few hundred meters away rather than the next street over. The existing literature on hot spots policing has focused mainly on catchment areas

whenever it fell within its geo-fence. Appendix B.1 reports further details. 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.

[21] Homicides are recorded by 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, and include relevant details such as location. Our endline survey (discussed below) suggests that official statistics record only about a fifth of all crimes.

[22] For details on all pre-specified aspects of the design see https://www.socialscienceregistry.org/trials/1156.

Figure 2: An example of assignment to the four treatment conditions



of about 1–2 blocks or about 100–150m.[23] We felt this radius could be too narrow, however, and opted for something more flexible.

Our preferred and pre-specified approach partitioned control segments into one of three experimental conditions according to their distance from the treated segment: <250m, 250–500m, and >500m. Figure 2 illustrates this partition, ignoring municipal services for simplicity. The segment at the center of the two radii was assigned to the intensive policing treatment. Nearby segments are classified by their distance to the treated segment.

This approach makes the estimation of treatment and spillover effects fairly simple: it is simply a matter of comparing weighted means of crime levels across these different treatment conditions. For instance, consider the case where we believe that spillovers do not extend beyond 250m. Then direct treatment effects are simply the difference in crime between directly treated segments and the subset of control segments more than 250m away from treated ones. Spillovers within the experimental sample are simply the difference between

[23]e.g. Braga et al. (1999); Braga and Bond (2008); Mazerolle et al. (2000); Taylor et al. (2011); Weisburd and Green (1995)

crime in segments 0–250m from a treated segment and those more than 250m away. We calculate spillovers into the non-experimental sample similarly. As we explain below, the density of treatment introduces some bias and hard-to-model clustering that requires additional corrections, but the basic principle of comparing means across treatment conditions is the core of our design.

Our approach ignores the possibility of spillovers beyond 500m, as well as non-spatial spillovers. Some crime is undoubtedly displaced in non-Euclidean ways (e.g., to possibly distant segments where the benefits of crime are high and the risk of detection is low).[24] Within the 500m radius, we also need a pre-specified rule for deciding whether to use the 250m or 500m radius for spillovers. We discuss this below.

4.3 Randomization procedures, design, and balance

We used a two-stage randomization procedure to maximize the spread between segments assigned to each experimental condition. This ensured as many segments as possible had a high probability of assignment to the 250–500m and >500m conditions. We first blocked our sample by the 19 police stations, then randomized segments to intensive policing in two stages: first assigning quadrants to treatment or control, then assigning segments within treatment quadrants. We assigned no more than two segments per quadrant to intensive policing. This procedure assigned 756 segments to intensive policing and 1,163 to control.[25]

In March 2016, we selected streets for municipal services. We sent enumerators to take five photographs and rate segments for the presence of disorder.[26] Of the 1,534 segments they were able to safely visit, 70% had at least one maintenance issue. We made these, plus the 385 segments they could not visit safely, eligible for municipal services assignment. We blocked on police station and the previous intensive policing assignment, and assigned 201 segments (14% of eligible segments) to municipal services.[27]

[24] Ferraz et al. (2016) find evidence of non-spatial spillovers in Rio de Janeiro's favela pacification. We expect these non-spatial spillovers could lead us to overstate direct treatment effects and understate total spillovers.

[25] Within each station we took all quadrants with at least one segment and randomized quadrants to treatment with 0.6 probability. We then used complete randomization to assign eligible segments to treatment within treatment quadrants.

[26] They looked for graffiti, garbage, and run-down buildings. A limitation is that we measured disorder after two months of policing treatment. We had no reason to expect the treatment to affect physical disorder, and there is no statistically significant difference between experimental and non-experimental segment.

[27] These 201 were the first "batch" to be treated. We also randomized a second batch of 214 segments 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 segments 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.

Table 1: Distribution of treatment and spillover assignments across the experimental sample

		Municipal services assignment to:					
		Treatment	<250m	250m-500m	>500m	*Ineligible*	All
		(1)	(2)	(3)	(4)	(5)	(6)
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 having no observed garbage or broken lights. For simplicity, we ignore whether ineligibles are <250m to intensive policing or municipal services segments or not.

Table 1 summarizes how the 1,919 experimental segments are distributed across 20 treatment conditions and potential outcomes—4×5 conditions tied to the four conditions for each intervention (treatment, <250m, 250-500m, and pure control) plus the ineligible category of streets that we deemed were in no need of municipal services.[28]

As described above, we can calculate treatment and spillover effects by comparing crime across these treatment conditions. Moreover, in the event we do not find any evidence of spillovers beyond 250m (as expected, and as demonstrated below) this design and pre-specified rules allow us to combine the 250-500m and >500m conditions into a single "control" condition, hence reducing the number of comparisons we make.

Table 2 reports summary statistics for the experimental sample. In October 2016, the police updated all 2012–16 crime data with more accurate GPS coordinates and additional crime categories, and we report both the original and updated data.[29] Experimental segments had between 0 and 82 crimes reported in the previous four years (461 with the updated data as we had information on more crime types), with an average of 5 reported crimes per segment.[30] More than half were property crimes, but violent crimes such as murders and assaults were also important. 95% of segments had relatively low levels of physical disorder such as garbage. We will return to crime levels in the following section, when we consider the survey data on unreported crimes.

Random assignment produced the expected degree of balance along covariates. Table 2 reports the weighted means for a selection of baseline covariates, by experimental assignment, for experimental and non-experimental segments. For the most part, background attributes

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

[29]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.

[30]Quadrants with at least one segment 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.

Table 2: Descriptive statistics for the experimental sample (N=1,919) and tests of balance (treatment versus all control streets, including potential spillover streets)

					WLS test of balance			
	Summary statistics				Intensive policing		Municipal services	
	Mean	Std. Dev.	Min.	Max.	Coeff.	p-val	Coeff.	p-val
Variable	(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)	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 segments in quadrant	127.21	86.99	2	672	2.05	0.71	-3.04	0.57
# of experimental segments in quadrant	3.67	2.68	1	14	-0.30	0.08	-0.16	0.31
# segments treated with policing in quadrant	1.15	0.95	0	3	1.35	0.00	-0.01	0.91
# segments treated with services in quadrant	0.66	0.69	0	3	-0.08	0.06	0.91	0.00
Assigned to intensive policing	0.48	0.50	0	1	1.00	-	0.00	-
<250m from intensive policing	0.29	0.46	0	1	-0.56	0.00	0.01	0.83
250–500m from intensive policing	0.14	0.35	0	1	-0.28	0.00	0.00	0.96
>500m from intensive poling	0.09	0.28	0	1	-0.17	0.00	-0.01	0.72
Assigned to municipal services	0.41	0.49	0	1	0.00	-	1.00	-
<250m from municipal services	0.19	0.39	0	1	0.05	0.01	-0.31	0.00
250–500m from municipal services	0.17	0.37	0	1	-0.01	0.71	-0.28	0.00
>500m from municipal services	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.

appear balanced across experimental conditions. There are some minor differences between treatment and control segments (for instance, treated segments are slightly less likely to be in industrial zones), but overall the imbalance is consistent with chance and is robust to alternative balance tests.[31]

4.4 Procedure for determining the relevant spillover radii

To determine the relevant spillover radii, we pre-specified a procedure: if there is no evidence of a statistically significant difference between the 250–500m and >500m regions using a $p < .1$ threshold, then we collapse them into a single control condition and the spillover condition will include streets <250m from treated segments only. Furthermore, if there is no statistically significant difference between streets in the <250m and the >250m control region using a $p < .1$ threshold, then we ignore spillovers altogether (i.e. ignore the partition of control streets into various conditions) and estimate the β coefficients alone in a simple treatment-control comparison.[32]

4.5 Estimation

With this design, we can estimate any treatment effect by comparing weighted average crime levels across the experimental conditions in Table 1. We use regression estimators to control for possible confounders, but the estimated treatment coefficients have the same interpretation—as mean differences. The regression specification is:

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

[31] 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 segments in the experimental sample (treatment, <250m, 250-500m and >500m), or the three-category assignments for streets in the non-experimental sample (<250m, 250-500m and >500m). To obtain exact p-values, we 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. The p-value is non-significant, as expected, for both the experimental and non-experimental samples: $p = 0.681$ for segments and $p = 0.531$ for non-experimental segments. We draw similar conclusions from tests of treated vs control units >250m away and between control units <250m and >250 away.

[32] In retrospect, this pre-specified rule was too permissive. 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 imprecisely-estimated spillovers with a $p > .1$ that are nonetheless large enough to offset any direct treatment effects. As we will see, this is not an issue in our case. The spillovers within 250m are economically significant in that they can more than outweigh the direct treatment effects, and some tests suggests they are significant at almost exactly the $p = .1$ level. In accordance with the pre-specified rule we account for these important spillovers. Nonetheless, slight changes could have compelled us to ignore spillovers in our main specifications. Hence in future, rules for flexible spillovers may want to be more permissive.

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 <250m or <500m from treatment, or a vector of both indicators); γ is a vector of police station fixed effects (our randomization strata); and X is a matrix of pre-specified baseline control variables.[33] Weights are the inverse probability weights (IPWs) of assignment to each experimental condition.

Whereas equation 1 estimates spillovers only within the experimental sample of 1,919, we can take advantage of the fact that tens of thousands of additional streets neighbor our treated segments, and estimate spillovers on the full range of streets by pooling the experimental and nonexperimental samples to run the following weighted least squares regression:

$$\begin{aligned} Y_{sqp} = \beta_1^P P_{sqp} + \beta_2^P M_{sqp} + &\beta_3^P (P \times M)_{sqp} + \lambda_1^P S_{sqp}^P + \lambda_2^P S_{sqp}^M \\ &+ \lambda_3^P (S^P \times S^M)_{sqp} + \tau E_{sqp} + \gamma_p^P + \Theta^P X_{sqp} + \delta^P (E \times X)_{sqp} + \epsilon_{sqp}^P \end{aligned} \quad (2)$$

where E_{sqp} is an indicator variable that takes the value of one for experimental street segments. For example, 77,848 nonexperimental segments lie within 250m of one of the 1,919 streets in the experimental sample, significantly improving power. Just as we partition the experimental control group into spillover and pure control conditions, we partition the nonexperimental sample in the same way. This pooled sample constrains the estimated λ coefficients to be the same for all spillover segments, regardless of whether they are in the experimental or nonexperimental sample.[34] This is our preferred estimation approach in the paper, with alternatives presented in the appendix.[35]

Each of these regressions preserve the comparison of means across treatment conditions.

[33] We selected these covariates by their ability to predict baseline crime levels. X also includes an indicator for segments ineligible for municipal services treatment by virtue of their baseline disorder.

[34] If, however, we do not want to pool the samples, it is possible to calculate nonexperimental spillovers through the weighted least squares regression on the 62,824 segments alone:

$$Y_{sqp} = \lambda_1^N S_{sqp}^P + \lambda_2^N S_{sqp}^M + \lambda_3^N (S^P \times S^M)_{sqp} + \gamma_p^N + \Theta^N X_{sqp} + \epsilon_{sqp}^N$$

using IPW for assignment to the conditions S^P and S^M.

[35] 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 2). 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 (1) and (2) maintain the spirit of the original estimation approach but correct for these problems.

In equations 1 and 2, the omitted condition is the control segments beyond a radius of either 250m or 500m, following the pre-specified rule above. The coefficients on treated and spillover conditions estimate crime differences relative to the control segments. In particular, β_1 and β_2 estimate the marginal intent-to-treat (ITT) effects of each treatment alone and β_3 estimates the marginal effect of receiving both. A negative sign on β_3 implies positive interactions or increasing returns. The effect of receiving both interventions is the sum, $\beta_1 + \beta_2 + \beta_3$.[36]

Why use inverse probability weights? Spillovers introduce spuriousness that can be corrected with IPWs. Experimental segments close to other experimental segments, 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 could mechanically lead us to conclude that there are adverse spillovers. Controlling for baseline characteristics and crime histories reduces but does 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.[37] These weights ensure that all segments have the same probability of being exposed to spillovers. As we will see, with baseline controls, the IPW correction does not make a major difference to our estimates. Nonetheless we include them for propriety's sake.

4.6 Alternative spillover estimation with continuous decay

Instead of partitioning control segments into bands, we could have assumed that spillovers follow a continuous, monotonic spatial decay function, and estimate direct and spillover effects with the following OLS regression:

[36] Because some streets were not eligible for municipal services, the sum of the three β estimates is not the exact estimate of receiving both interventions. However, the difference is trivial and we opt for this estimation of combined effects for simplicity. Since every street is not eligible for all three treatment combinations (because of the eligibility for the municipal services treatment), when we add up the three β coefficients, we are pooling effects over different subgroups whose effects could be heterogeneous. This implies we are effectively constraining the ITTs to be the same for the same treatment condition across eligibility strata. Since we control for these eligibility strata, we ensure we are not confounding treatments with unobservables. These controls come in the form of IPWs for the treatments and the dummy for the eligibility strata included in X. In any case, since we do not find strong evidence of treatment effects, there should be less concern on the presence of heterogeneous effects over the different subgroups.

[37] Each segment's probability of exposure to <250m or 250-500m spillovers can be estimated with high precision by simulating the randomization procedure a large number of times. Such IPWs 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 B.2 describes and maps IPWs in our sample.

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

where $f(d_{sqp,t})$ is a spatial decay function with a standardized distribution. This function is a weighted sum of distances to all treated segments, where t enumerates treated segments and T_P and T_M are the set of all treated segments with intensive policing and municipal services, respectively. Treated segments receive no spillover from themselves but can receive spillovers from other treated segments. Applied to the non-experimental sample, the regression omits direct treatment effects.

We consider an exponential decay function, $f(d_{sqp,t}) = 1/(e^{d_{sqp,t}})$, as well as a simple inverse linear decay. We can no longer employ IPWs to weight street segments because the exposure measures are continuous variables. Instead, we include in the control vector the expected spillover intensities (averaged across 1,000 simulated random assignments) and the probabilities of being treated by each intervention. We calculate statistical significance using randomization inference.[38]

4.7 Why randomization inference?

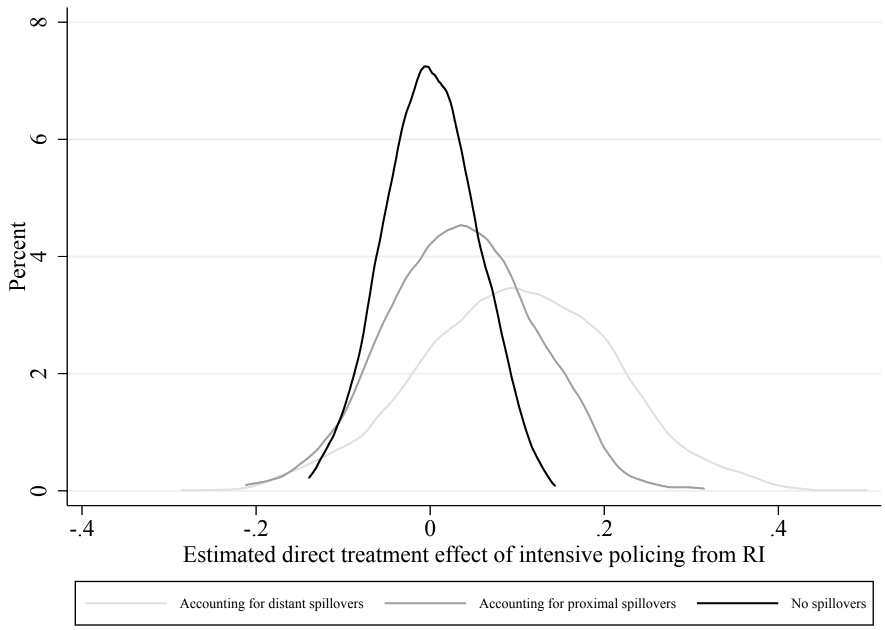
Randomization inference (RI) gives precise p-values based on the empirical distribution of all estimated treatment effects that could arise under our design and data under the null hypothesis of no effect for any unit. RI reassigns treatment randomly thousands of times, each time estimating the treatment effect that could have arisen by chance from that comparison. Figure 3 displays the empirical distributions of estimated direct treatment effects for intensive policing under three variants of the estimating equation in (1): the simple no-spillovers case (i.e. $S_s^P = S_s^M = 0$ for all s); the case where S^P and S^M indicate spillovers within 250m only; and the case where S^P and S^M indicate spillovers within 500m.

Most importantly, the distribution widens when accounting for spillovers. The no-spillovers case has the narrowest distribution. The distribution widens as we account for wider spillover regions. That is, we are more likely to get large treatment effects by chance.

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 shrinks dramatically and begins to exclude high-crime regions of the city. Experimental segments that are close to other experimental segments are assigned to the spillover condition in most randomizations, creating patterns of "fuzzy clustering" (Abadie et al., 2016). These

[38]For each of 1,000 simulated random assignments we obtain a simulated ATE. The standard deviation from this empirical distribution of ATEs is the standard error of the estimates.

Figure 3: The empirical distribution of estimated treatment effects on insecurity under different spillover scenarios



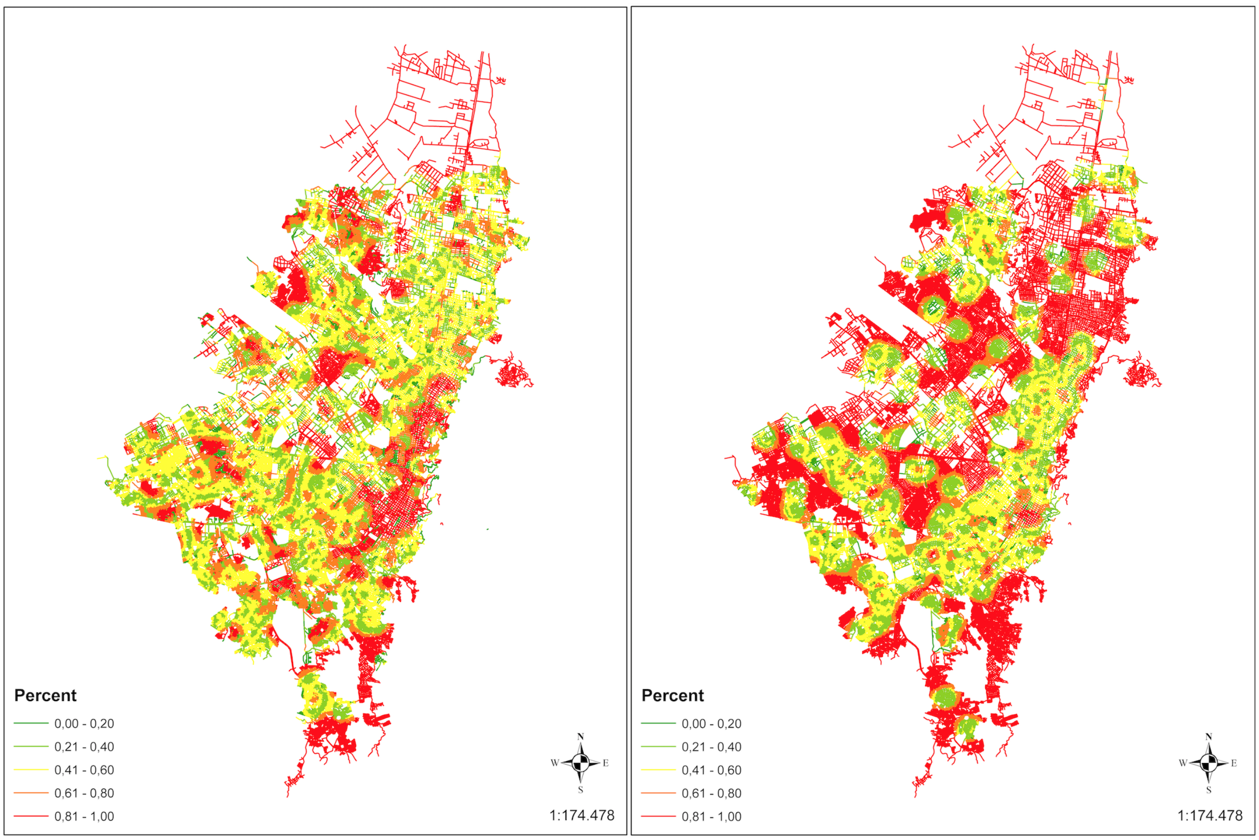
Notes: The figure displays the empirical distribution of treatment effects on the insecurity index for intensive policing. We simulate the randomization procedure 1,000 times and estimate treatment effects for each randomization using post-treatment data under the sharp null of no treatment effect for 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 250m; and the case where S^P and S^M indicate the larger spillover area within 500m.

clusters are difficult to model because they have to do with distance from other experimental segments rather than an observed characteristic such as a quadrant.

We can see the fuzzy clustering in Figure 4, which illustrates for each segment the proportion of segments within 500m that are assigned to the same experimental condition, including the spillover conditions. For instance, for intensive policing, most segments in the dense city center (the middle right of the map) have neighbors with the same high probability of assignment to the <250m spillover condition. For municipal services, there are large swathes of the city with a high probability of assignment to the control condition, forming a cluster that does not conform to administrative boundaries. The figures imply that, instead of having thousands of independent segments, we actually have dozens of clusters. But there is no geographic marker for them.

Finally, the simulations in Figure 3 show that the distributions of simulated treatment effects with spillovers are not centered at zero. 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 potential outcomes. When we ignore spillovers, we stipulate that there is no such clustering, which

Figure 4: Fuzzy clustering in the presence of spillovers



Notes: The figure displays the proportion of segments within 500m assigned to the same treatment condition for intensive policing (left) and municipal services (right).

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). Unfortunately, the spillover 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 estimated effects from the 10,000 randomizations. We use these RI p-values in place of the conventional standard errors-based p-values whenever we estimate treatment effects in the presence of spillovers. Additionally, the simulations used in the RI procedure provide an estimate of the bias. All of our tables report bias-corrected treatment effects. Appendix B.2 reports the specific biases estimated.

5 Data

We draw on six main sources of data.

1. *Administrative data on police and municipal services compliance.* The police shared the full database of GPS patrol locations for all 136,984 streets, 2015–17.[39] 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. *Officially reported crimes and calls for service.* Police shared data on reported crimes and operations 2012–17, geolocated to 136,984 streets.[40]

3. *Crime survey of 24,000 Bogotá residents.* We complemented administrative data with a survey for three reasons. First, as we will see below, a majority of crime and nuisances go unreported. This is one reason that police identified as "hot" many streets that had

[39] 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 33 of the 37 weeks of the intervention.

[40] Prior to the intervention, we received the 2012–2015 data on the city's priority crimes: homicides, assaults, robberies, and car and motorbike theft. 77% of the crimes had exact coordinates and the rest had the address, which we geolocated ourselves, with about 71% success (or 93% of all reported crimes). 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. Some U.S. studies use emergency call data. Initially these were not available, and our pre-analysis plan excluded them. Later, partially complete data became unexpectedly available, and our main results are robust to their inclusion (not shown).

no officially reported crimes. Second, we were concerned that treatment could increase crime reports, thus inflating our treatment effects, and wanted to be able to test this concern. Third, we wanted to measure outcomes such as citizen trust in the state and police.

In October 2016 we surveyed approximately 10 people per street segment on 2,399 segments—the 1,919 in the experimental sample, plus a representative sample of 480 non-experimental segments. We interviewed a convenience sample and average responses over each segment. The survey collected outcomes such as: perceptions of security risks; perceived incidence of crimes; crimes personally experienced; crime reporting; and trust in and perceived legitimacy of the police and the Mayor's office.

4. *Survey of street disorder.* To measure levels of street disorder before and after treatment, we sent enumerators to take photographs and rate the presence of graffiti, garbage, and boarded-up buildings on a 0–5 scale.[41]

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 (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 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 police behavior. They also interviewed citizens in each segment about police behavior and attitudes.

5.1 Outcomes

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

[41] 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 experimental segments. We did not collect data in the remaining 385 streets because of security concerns from the enumerators. (Note that there was no association between intensive policing treatment and these security concerns.) As we discuss in section 4.3, 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.

indices to reduce the number of hypotheses tested (following Kling et al. 2007).

Our primary outcome is crime. To measure aggregate and spillover effects, there is only one measure available for all streets: officially reported crimes on the segment. But to measure direct effects, we prespecified two insecurity measures (one of which includes officially reported crimes):

1. *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 an 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.

2. *Crime incidence on the segment.* We construct a standardized index of crime that equally weight: (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;[42] and, (ii) officially-reported crime incidents on that segment since the beginning of the intervention. We can subdivide all measures into property and violent crimes, although our main measure pools all crimes into one index.

We discuss secondary outcomes, particularly the perceived legitimacy of the police and local government, in Section 7.6 below.

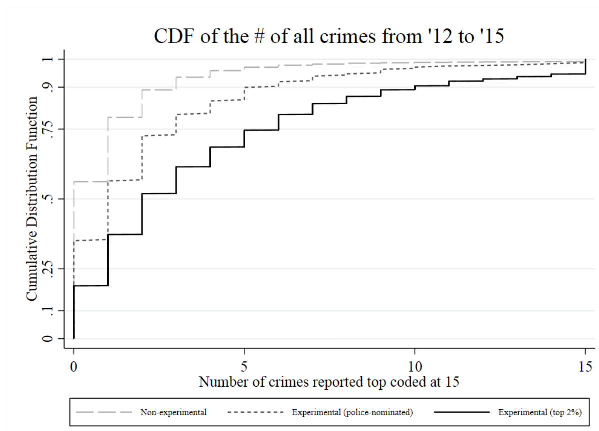
5.2 Insecurity in the experimental and non-experimental samples

Our experimental sample includes a range of segments from moderate to very high levels of crime. For instance, Figure 5 displays cumulative distribution functions (CDFs) for officially reported crimes in the pre-intervention period 2012–15 (Panel a), and during the 8 months of the intervention (Panel b). We plot three CDFs per panel: (i) the 135,065 non-experimental segments; (ii) the 248 experimental segments nominated by the police; and (iii) the 1,671 that were in the top 2% of reported crime.

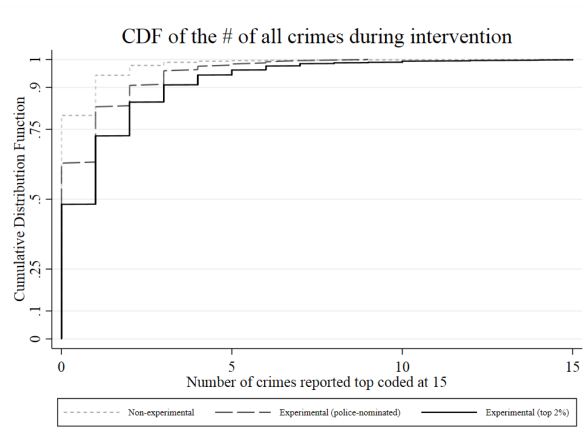
[42]The survey measured perceived incidence and personal victimization by walking respondents through a list of 11 criminal activities. After finding out whether any of these activities happened on the street since the beginning of the year, we asked respondents about each crime to establish perceived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

Figure 5: Cumulative distribution functions (CDFs) of crime measures, by experimental and non-experimental samples

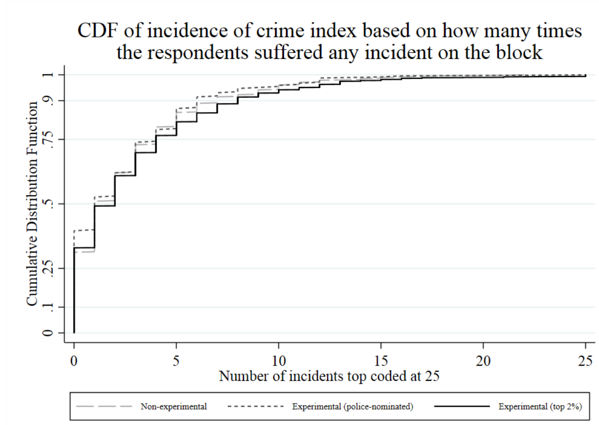
(a) CDFs for the pre-intervention period



(b) CDFs for the intervention period



(c) CDFs of incidence of crime index



(d) CDFs of perceived crime index

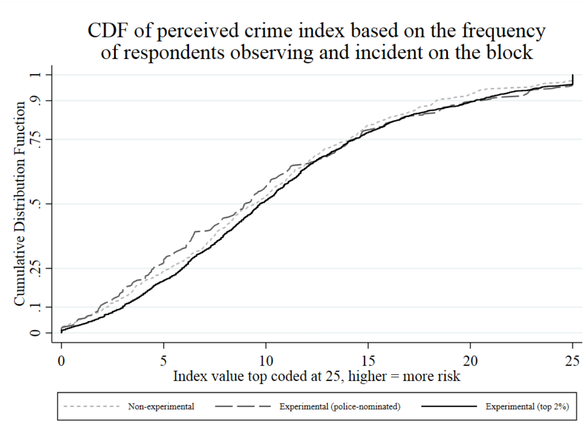


Table 3: Summary statistics of crime in the experimental and non-experimental samples (administrative and survey measures)

Dependent variable		Non-exp	Experimental sample, by source: Police	Experimental sample, by source: "Top 2%"		t-test of difference in means: Non-exp vs Police	t-test of difference in means: Non-exp vs Top 2%	t-test of difference in means: Police vs Top 2%
# of officially reported crimes pre- and post-treatment (2012–16)	Mean	1.685	4.763	7.621	Diff.	3.078	5.935	2.857
	SD	2.827	17.380	24.821	SE	1.118	0.621	1.266
	N	480	245	1,671	p-value	0.006	0.000	0.025
# of reported crimes pre-treatment (2012–15)	Mean	1.352	4.024	6.393	Diff.	2.672	5.041	2.369
	SD	2.541	16.885	23.804	SE	1.085	0.594	1.226
	N	480	248	1,671	p-value	0.015	0.000	0.054
# of reported crimes during intervention	Mean	0.333	0.739	1.227	Diff.	0.405	0.894	0.489
	SD	0.729	1.320	2.140	SE	0.091	0.062	0.099
	N	480	245	1,671	p-value	0.000	0.000	0.000
Index of survey-based crime, post-treatment, z-score	Mean	-0.079	-0.090	0.036	Diff.	-0.011	0.115	0.126
	SD	0.839	0.920	1.051	SE	0.070	0.046	0.064
	N	480	245	1,671	p-value	0.879	0.013	0.051
Perceived incidence score (0-60), higher = more risk	Mean	10.104	10.196	10.728	Diff.	0.092	0.625	0.532
	SD	6.438	7.203	6.627	SE	0.546	0.336	0.488
	N	480	245	1,671	p-value	0.866	0.063	0.276
# of all crimes respondents personally experienced	Mean	2.671	2.510	3.138	Diff.	-0.161	0.467	0.627
	SD	3.435	3.656	5.779	SE	0.281	0.211	0.273
	N	480	245	1,671	p-value	0.568	0.027	0.022

Table 3 reports means and mean differences between the three samples. By construction, reported crimes are greatest in the "top 2%" sample and next highest in the police-selected sample. Reported crimes are lowest in the non-experimental sample, as expected. On average, streets in the top 2% have about 5 times as many reported crimes as those in the non-experimental sample.[43]

Nonetheless, we can see from the CDFs that a number of police-nominated and even some top 2% streets have just 0–2 crimes in the pre-intervention period. Also, a small number of the non-experimental streets have a sizable number of crimes. Why is this so? High-crime non-experimental streets are simple to explain: the police limited the number of treated streets to two per quadrant. In high-crime quadrants, this means many relatively high-crime segments are in the non-experimental sample. To explain the low-crime "top 2%" streets, above we noted that in 2016 the police issued a more complete and correct version of their 2012–15 geo-located crime data. Some "top 2%" segments had some crimes reclassified away from them but remain in the sample nonetheless.[44]

Remember, however, that none of the experimental streets should be truly "low-crime", even if there is no officially reported crime. Local police stations reviewed every candidate for the experimental sample and threw out those that were low in crime or nuisances. Those with low levels of reported crime presumably have high levels of unreported crime and nuisances.

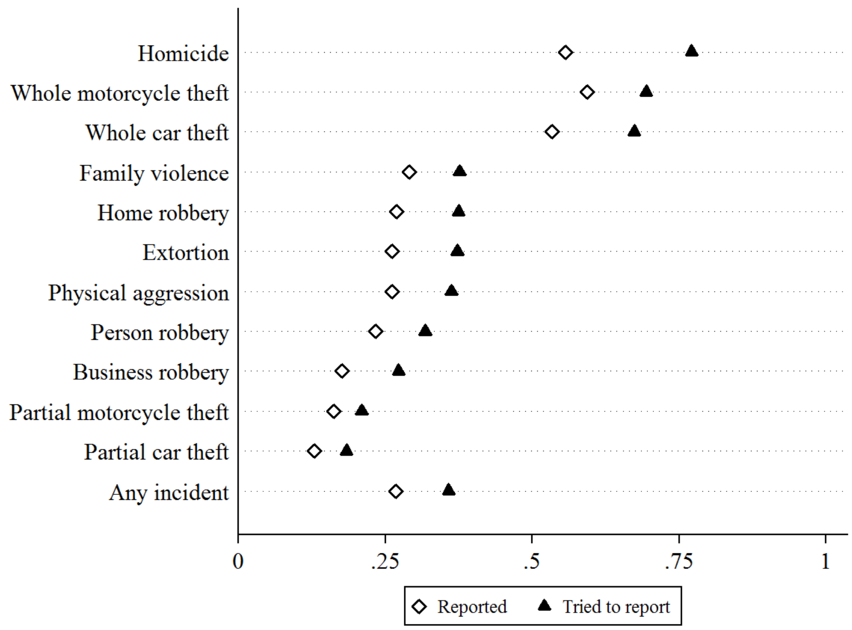
Figure 6 illustrates the difference between actual and officially-reported crimes. For 11 crimes, the survey asked whether or not people had experienced a crime since the beginning of the year, whether they had attempted to report it, and if they were successful. Homicides are reported by police if individuals did not report them, so administrative data probably capture most or even all murders. But for the other 10 crimes, 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. Reporting rates are highest for vehicle theft, because insurance claims require a report. Otherwise the vast majority of crimes are never reported.[45]

[43]Appendix Table C.1 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.

[44]Other reasons include the fact that less serious crimes were given less weight in the sample selection, so that a street with one murder was more likely to enter the experimental sample than one with several muggings. Finally, a small miscalculation in the sample selection admitted a small number of moderate crime streets into the "top 2%" sample.

[45]These survey data also provide an opportunity to test whether people were more likely to report crimes to the police on treated segments. If so, this would call into question any treatment effects based on reported crime data. We see no difference in the survey-based likelihood of crime reporting on treated streets. 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. This suggests that administrative data are suitable for outcome assessment even while the treatment is being delivered.

Figure 6: 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.

Panels c and d of Figure 5 display CDFs for two composite survey measures: an index of 10 crimes (excluding homicide), where respondents were asked for each crime to rank their perceived frequency of the crimes on that segment on a 0-6 scale; and for the same 10 crimes an additive index of the number of times each of the 10 respondents personally experienced one of these crimes in the previous eight months (the treatment period). Because these are post-treatment measures, we have to take the difference between the samples with caution. Treatment and spillover effects would likely bias the three samples to be closer to one another.

By these metrics, our experimental sample has moderate levels of crime on average. For instance, 3 in 10 of the people stopped on each of the experimental streets reported a personal experience of crime on that segment in the previous 8 months. This is a relatively high rate of victimization. Perceived risk is 10 of 60 on average, stretching as high as 20 or 30 in the highest crime streets.

What this means is that only a fraction of our experimental sample is comparable to what the U.S. literature calls a "hot spot." This is an unavoidable consequence of scale. As a result, we should compare effectiveness to other hot spot interventions with caution. An advantage of the larger and broader sample, however, is that we can estimate the effect of increased state presence on crime in a mostly normal set of streets. We can then "simulate" a hot spots intervention by looking at impacts on the subsample of highest crime streets.

6 Program implementation and compliance

The police and municipal services agencies largely complied with 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 each program on various first-stage outcomes. We estimate equation (1) ignoring interactions between the two treatments and spillovers, since we have no expectation of either in this first stage.

Patrol time Our main measure of policing is average patrol minutes per day on each segment. We estimate control streets received 92 minutes of patrolling time per day, on average. Treated streets received an extra 77 minutes, an 84% increase. By comparison, non-experimental received an average of 33 minutes of patrolling time per day.[46]

Our best assessment is that the increase in patrol time on treated streets did not take a material amount of time away from control segments, for two reasons. First, there are 130 segments in the average quadrant, and so the 77 minute rise on two segments means just a minute less time for all other segments. Second, the introduction of the patrol geo-locators was designed to increase the efficiency and time on the street of patrols, and our best assessment is that all segments received at least 10–20% more patrol time than the pre-intervention period.[47]

Police actions We see no effect of increased policing on arrests or police actions such as drug seizures. This implies any direct effect of the policing comes from deterring or displacing criminals rather than incapacitating them. Incapacitation, of course, would reduce the chance that crimes are displaced.

Services The evidence on service delivery compliance is more mixed. Table 5 summarizes municipal services compliance. After assigning 201 segments to municipal services, city agencies diagnosed each one in March. They identified 123 segments needing clean-up services, and 47 needing lighting improvements. They performed the services June through August. Tree pruning and graffiti cleaning were one-time treatments; rubbish collection was expected

[46] Naturally, the devices that track patrol locations every 30 seconds periodically malfunction, and occasionally the system has an outage. Thus any estimate of minutes is probably an underestimate, one that is unlikely to be correlated with treatment.

[47] The survey asked whether citizens noticed an increase in patrols in the previous 6 months. On control segments, 13% reported an increase, compared to 21% on treatment segments.

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

Dependent variable	Control mean	ITT and standard error of assignment to:			
		Intensive policing		Municipal services	
	(1)	(2)	(3)	(4)	(5)
A. Intensive policing measures:					
Proportion of respondents who say police presence increased in last 6 mo.	0.129	0.076	[.011]***	0.017	[.013]
Daily average patrolling time, excluding quadrant-days without data	92.001	76.571	[4.424]***	-3.333	[4.371]
# of arrests	0.333	-0.053	[.082]	0.026	[.102]
# of drug seizure cases	0.041	-0.002	[.020]	0.029	[.024]
# of gun seizure cases	0.009	0.006	[.008]	0.007	[.013]
# of recovered car cases	0.003	0.000	[.001]	-0.003	[.001]*
# of recovered motorbike cases	0.006	-0.028	[.019]	0.032	[.027]
B. Municipal services implementation measures					
Proportion of respondents who say municipal presence increased in last 6 mo.	0.144	0.005	[.010]	0.016	[.012]
City determined segment is eligible for lights intervention	0.349	-0.007	[.048]	-0.139	[.048]***
Received lights intervention	0.000	-0.010	[.020]	0.199	[.026]***
City determined segment is eligible for garbage intervention	0.000	0.011	[.025]	0.627	[.032]***
Received garbage intervention	0.000	0.015	[.026]	0.382	[.033]***
June 2016 enumerator assessment of street conditions:					
Graffiti on segment	0.749	-0.018	[.050]	0.077	[.043]*
Garbage on segment	0.251	0.071	[.061]	0.015	[.049]
Visibly broken street light on block	0.000	0.012	[.012]	0.008	[.008]
December 2016 enumerator assessment of street conditions:					
Graffiti on segment	0.624	0.019	[.053]	0.059	[.047]
Garbage on segment	0.245	0.021	[.051]	0.002	[.043]
Visibly broken street light on block	0.029	0.022	[.016]	-0.015	[.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	41 (87.2%)
cleanliness	Ineligible for garbage	26	52	78	
assessment	All	47	154	201	
% of eligible streets receiving clean-up		74 (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.

to be semi-regular. Based on city data, 74 of the 123 streets (60%) were cleaned up, and in 41 of the 47 streets (87%) they repaired broken lights and replaced poor lights with better ones. No graffiti was cleaned-up.

The impacts were not obvious to residents. About 14% of survey respondents on control segments noticed an improvement in service delivery in the past six months, and this was only 1.6 percentage points greater in treatment streets (not statistically significant, see Table 4). We also visited segments in daytime in June and December 2016 to photograph and rate the streets. The before and after photos generally display relatively tidy streets and before-after differences are imperceptible. It is possible that lights repairs were more evident, but it was unsafe to visit segments at night. We see no effect of treatment in Table 4. One possibility is that the extensive margin is the wrong margin to evaluate, and another is that the disorder in cleaned up segments could have re-accumulated over days or weeks.

7 Program impacts

To assess direct, spillover, and aggregate impacts, we start by estimating program impacts on officially reported crime using all city streets. We then look more deeply at direct program effects using the survey and administrative measures of insecurity using only the streets with survey data.

In all these analyses, we estimate equation (2) above. Unless otherwise noted, our spillover condition is limited to streets within 250m of treated segments only. This follows from our pre-specified rule, as we do not see a statistically significant difference in crime between streets in the 250–500m and >500m regions, but we do see a difference between those <250m and >250 away. Appendix C.2 reports this spillover analysis.

7.1 Program impacts on officially reported crime

Table 6 reports estimates of treatment and spillover effects on officially reported crimes. The table reports impacts with and without the the interaction term between intensive policing and municipal services.[48]

Direct treatment effects We see only weak evidence that increases in police patrols and municipal services reduce crime. If we ignore the interaction between police and municipal services (columns 1–4 of Table 6), both intensive policing and municipal services reduce officially reported crimes slightly, but the coefficients are not statistically significant. Control segments report 0.743 crimes on average over eight months of intervention. Intensive policing reduced this by -0.099, a 13% improvement.[49] Municipal services reduced this by -0.133 crimes, an 18% improvement.

Once we include the interaction term in our estimating equation (columns 5–8) we see the largest and most statistically significant impacts in the segments that were assigned to both interventions. The coefficients on policing and municipal services alone actually switch signs to point to a slight increase in crime, although both are highly imprecise.

As a result, we see no evidence that either intervention on its own reduced crime. Rather, the decreases in crime appear to be concentrated in the segments that received both interventions. The coefficient on the interaction is -0.53, with an RI p-value of 0.010. The sum of the three coefficients is -0.423 with a p-value of 0.008 (not reported in the table). This sum corresponds to a 57% decrease in reported crimes on the 75 streets that received both interventions.[50]

[48] We omit the 57,695 streets with zero probability of assignment to the spillover condition. There are 51,390 non-experimental segments and 705 control segments for the policing intervention and 20,740 non-experimental segments and 546 control segments for municipal services. Thus even small estimated spillovers can have a large effect on the total crime estimates.

We pre-specified a one-tailed test 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 throughout. As noted above (footnote 35) we also pre-specified a pairwise analysis for treatment effects. While this proved to be an erroneous choice, Appendix B.3 reports those pre-specified pairwise results.

Appendix C.3 estimates the "unpooled" results on the experimental and non-experimental samples separately.

[49] We can see that these results are not driven by a decrease in patrolling time in control streets by estimating the marginal effect of one additional hour of patrolling time. The marginal effect of an additional hour of police patrols is a decrease of about 0.1 crimes. This is similar to the average effect of -0.099, as the average treatment street received 76 minutes of additional patrolling time. See Appendix C.8 for these IV estimates.

[50] Strictly speaking, we cannot simply add the three coefficients because not every street was eligible for municipal services. Because the estimated impacts of municipal services and both interventions are based on a subpopulation, it is technically incorrect to add the coefficient on intensive policing from the full sample. The estimated treatment effect of policing is almost identical whether we look at the full sample or the sample of municipal services eligible. And so we use the sum of the three listed coefficients for simplicity.

Table 6: Estimated direct, spillover, and aggregate impacts of the interventions, accounting for spillovers within <250m, pooling the experimental and non-experimental samples

	Dependent variable: # of 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	-0.099	*0.386*	756	-74.8	0.035	*0.664*	756	26.5
Municipal services	-0.133	*0.182*	201	-26.7	0.072	*0.597*	201	14.5
Both					-0.530	***0.010***	75	-39.8
Subtotal				-101.6				1.2
B. Spillover effect								
Intensive policing	0.017	*0.112*	52095	885.6	0.019	*0.108*	52095	989.8
Municipal services	0.002	*0.642*	21286	42.6	0.006	*0.967*	21286	127.7
Both					-0.007	*0.557*	15772	-110.4
Subtotal				928.2				1,007.1
Net increase (decrease) in crime				826.6				1,008.3
			95% CI	(-745, 2050)			*95% CI*	(-739, 2584)
			90% CI	(-463, 1874)			*90% CI*	(-497, 2296)

Notes: Columns 1–4 refer to the non-interacted results (equation 2 under the constraint that $\beta_3 = 0$ and $\lambda_3 = 0$) while columns 5–8 refer to the interacted results (equation 2 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.

The fact that the coefficient on the interaction is large, negative, and statistically significant (implying a relatively large reduction in crime) could imply that there are increasing returns to policing and municipal services, either because any increase in state presence has increasing returns, or because the combination of policing and services is somehow important.

We can multiply each estimated treatment effect by the number of treated streets to estimate the total direct effects of the two interventions. The total amount of crime directly deterred is small. Reallocating police and municipal services to higher-crime streets directly deterred just 101.6 crimes over eight months ignoring the interaction term. They did not deter any crime when we account for the interaction term. Neither total is statistically significant, however.

Spillover effects The evidence also suggests that any crime deterred may have been displaced to nearby segments. For intensive policing, all spillover coefficients in Table 6 are positive (including the sum of the three coefficients in Column 5), implying an increase in nearby crime. The p-values on the adverse spillovers from intensive policing are 0.112 without the interaction and 0.108 with the interaction.

There are so many segments (from the nonexperimental sample in particular) that these small spillover coefficients add up to high levels of displaced crime—1,007 when we allow for the interaction and 928 when we do not.

Aggregate effects We use these estimates to roughly assess the aggregate effect on crime city-wide. We estimate the total number of deterred crimes as the product of (i) the estimated coefficients and (ii) the number of treatment and spillover segments in the city. We then total all direct and spillover effects at the base of the table, and calculate RI confidence intervals for these totals.

These aggregate direct and spillover estimates suggest the treatments increased crimes by about 827 to 1,008 city-wide, or 2–3% relative to the total number of reported crimes. We have to take this increase with caution, as the estimates are not statistically significant at the 10% level. Moreover, this estimate would not capture general equilibrium effects if they exist (e.g. if the intervention is disrupting city-wide criminal networks).

While we cannot exclude zero spillovers, it is incorrect to view these aggregate estimates as imprecise. First, using the 90% confidence intervals we can rule out a decrease in city-wide crime of more than 1–2%. Second, recall that we were ex ante powered to detect spillovers of roughly 0.02 standard deviations–an order of magnitude more power than prior studies. Most of these spillover coefficients are just below that threshold. This is one reason why the confidence intervals on spillovers and aggregate effects include zero.

How does this compare to the spillover effects estimated in the systematic reviews? In a recent meta-analysis, the average point estimate for intensive policing was -0.104 standard deviations.[51] Our 90% confidence interval for the spillover effects of intensive policing on our insecurity index ranges from -0.110 to 0.124, meaning that the US mean is within but at the extreme tail end of our range.

Finally, as a thought experiment, we can use the coefficients in Table 6 to crudely estimate the aggregate effects of the program had the government delivered both interventions to all 882 treated streets (instead of just 75). We do so in Appendix C.4. We estimate this would have led to a fall of 373 crimes on directly treated streets, but this would have been outweighed by spillovers into experimental and non-experimental segments, for a net aggregate increase of 664 crimes.

Disentangling municipal services Our qualitative work and compliance data hinted that the lighting intervention may have been more compliant, effective, and persistent than the street clean-up. But the data do not support this conclusion. Both lighting and cleanup services appear to have been important. For example, we see no evidence that municipal services treatment effects were concentrated in the segments diagnosed as needing improved lights. Furthermore, we do not see larger treatment effects at nighttime (tables not shown).

7.2 Program impacts using the combined survey and administrative measures

The survey data tell a similar story as the administrative data on reported crimes. Table 7 reports impacts on our broader security measures for 1,916 experimental streets and 480 non-experimental streets with survey data. These measures include: the perceived risk index, based on surveys; and the index of crime, which averages survey- and officially-reported crime (with components displayed). (Note that the reported crime data is only for the 2,396 segments where we have survey data, and so differs from the previous table.) We also consider an average of the two measures, called the "insecurity index." Treatment effects can be interpreted as average standard deviation changes in the outcome, unless specified otherwise. The table includes the interaction between policing and municipal services (and omitting this interaction term does not materially change our conclusions, as shown in Appendix C.5).

Alone, each intervention is associated with a roughly 0.1 standard deviation security improvement on directly treated streets, not statistically significant. As with administrative

[51] See Braga et al. (2014). They report a positive coefficient, which in our context implies a negative sign (a reduction in crime). We switch the sign for convenience.

Table 7: Program impacts on security in the experimental sample, accounting for spillovers within 250m, with p-values from randomization inference, with interaction between treatments (N=2,396)

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interven-tions	Sum of (2), (3), and (4)	Impact of spillovers <250m: Any intensive policing	Any municipal services	Both interven-tions	Sum of (6), (7), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Insecurity index, z-score (+ more insecure)	-0.003	-0.098	-0.105	-0.125	-0.329	0.052	0.129	-0.215	-0.035
		0.386	*0.399*	*0.445*	***0.057***	*0.281*	***0.035***	***0.020***	*0.663*
Perceived risk index, z-score (+ riskier)	0.049	-0.116	-0.093	-0.090	-0.299	0.002	0.055	-0.153	-0.095
		0.244	*0.381*	*0.625*	***0.060***	*0.500*	*0.214*	***0.074***	*0.950*
Crime index, z-score (+ more crime)	-0.054	-0.048	-0.083	-0.119	-0.249	0.084	0.160	-0.206	0.038
		0.714	*0.591*	*0.422*	*0.163*	*0.175*	***0.011***	***0.032***	*0.360*
Perceived & actual incidence of crime, z-score (survey)	0.059	-0.072	-0.157	0.067	-0.163	0.045	0.101	-0.145	0.001
		0.519	*0.121*	*0.409*	*0.504*	*0.324*	***0.082***	*0.100*	*0.548*
# crimes reported to police on street segment (admin)	0.743	0.011	0.080	-0.442	-0.351	0.137	0.231	-0.270	0.098
		0.816	*0.406*	***0.038***	***0.085***	*0.283*	***0.023***	*0.134*	*0.288*

crime, we see the largest and most statistically significant impacts of state presence in the segments that received both interventions. Those 75 segments reported a 0.329 standard deviation decrease in the insecurity index, significant at the 10% level (column 5).[52] The coefficients on perceived risk and crime indexes are similar, though the combined effect of the two treatments is statistically significant only for perceived security.

Again, there is suggestive evidence of increasing returns to state presence, although it is less robust than with reported crimes alone. The coefficient on the interaction term (column 4) is statistically significant for officially reported crimes only.

The broader security measure also indicates that crime displaces to nearby control segments. Columns 6 to 9 of Table 7 report spillover effects. Intensive policing alone and municipal services alone are associated with increases in crimes on nearby segments of 0.5 to 0.13 standard deviations. Only the municipal services spillovers are statistically significant. The interaction terms are generally negative (see column 8) and generally statistically significant. Thus adverse spillovers appear to be strongest in the streets that received one intervention or the other.

7.3 Heterogeneity by type of crime

Police tend to prioritize violent crimes such as assault, rape and murder over property crimes such as burglary or theft. Table 8 takes the aggregate impacts on officially-reported crime from Table 6 and disaggregates these total effects into violent and property crimes.

The interventions have opposing effects on property and violent crime. Our best estimate is that aggregate violent crimes fell by 369 crimes in total (3% relative to the total number of violent crimes) when we account for the interaction between treatments. The two most socially costly crimes, homicides and sexual assaults, fall by 86. This represents a large proportion of very serious crimes—7% relative to the total number of homicides and sexual assaults citywide—even if the result is statistically not significant. Neither decline is statistically significant at the 10% level, though it is almost so. Given the gravity of these crimes, we should not dismiss these decreases, however imprecise.

Property crimes rose by 1384 in aggregate (5% relative to the total number of property crimes). This increase is statistically significant at the 10% level when we include the interaction. The difference between aggregate effects in property and violent crimes is also

[52] After completion of the experiment, we also received calls-for-service data from police. We did not prespecify that we would use these administrative data. Also, we are concerned that direct treatment would directly affect calls for service, especially the more frequent presence of police. Hence we omit these data from the final analysis. The average experimental segment received 17.5 calls over the eight months. Intensive policing alone reduced this by 3.9 calls (p=0.30), municipal services increased calls by 1.7 (p=0.44), and the cumulative effect of both interventions was to reduce calls by 2.3 (p=0.71).

Table 8: 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	812.9	(-992, 1826)	(-676, 1632)	1015.0	(-742, 2582)	(-500, 2297)
Property crime	989.7	(-438, 1813)	(-282, 1646)	1384.0	(-122, 2681)	(153, 2453)
Violent crime	-176.8	(-874, 378)	(-759, 280)	-369.0	(-1086, 287)	(-955, 161)
Homicides and sexual assaults	-59.6	(-182, 52)	(-165, 39)	-86.0	(-227, 48)	(-197, 27)
Difference between property and violent crime	1166.5			1752.9		
p-value	0.063			0.000		

Notes: This table presents the aggregate effect calculation for various crime subgroups assuming spillovers within 250m. Calculations are based on the aggregate effect and confidence interval described in Table 6.

statistically significant at the <1% level with the interaction between treatments.

Table 9 details the direct, spillover and aggregate impacts for violent crimes only. (Appendix C.6 reports more detailed impacts for property crime). We should take these subgroup estimates with some caution, as we have not adjusted the p-values for multiple hypothesis testing. Nonetheless, note that intensive policing has a statistically significant direct impact on violent crime when not considering the interaction. The estimates generally suggest there are beneficial spillovers on violent crime.

7.4 Heterogeneity by level of initial crime

Heterogeneity in direct effects We pre-specified one major form of heterogeneity analysis, by baseline levels of crime. This helps us to compare our experimental results to the U.S. hot spot policing literature. Broadly speaking, we observe substantially larger treatment effects on the 10–20% highest crime streets, but below that the treatment effects are relatively similar.

Figure 7 reports estimated direct treatment effects on the insecurity index for the n% highest-crime hot spots. Specifically, we estimate equation 2 nine additional times. Each time, we interact each treatment indicator with an indicator for whether a segment is above the nth percentile of baseline crime levels among our experimental sample of hot spots, for $n = 0, 10, 20, ..., 90$. The figure plots the coefficients on these higher crime streets, with the n% "hottest spots" on the right.

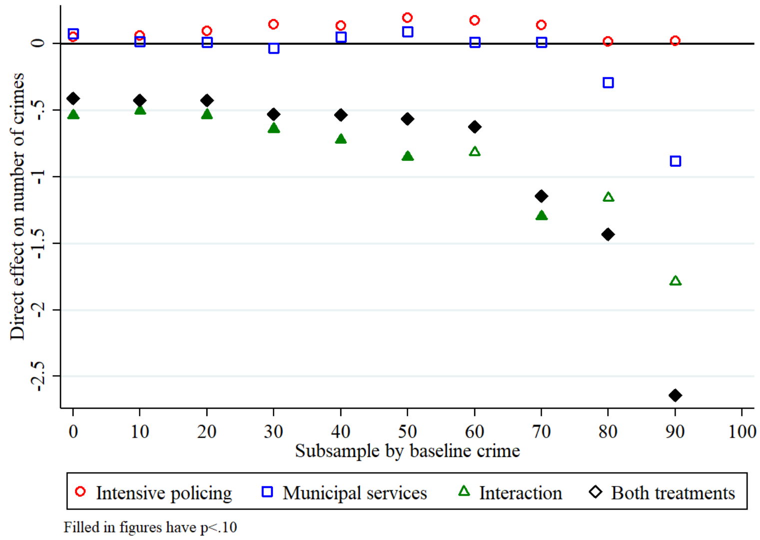
The direct treatment effect is fairly constant up until the point we reach the street segments in the 70^{th} percentile and above. After this point the impact of receiving both interventions climbs first to 1 fewer crime and then to 2.5 fewer crimes, on average, during

Table 9: Estimated direct, spillover, and aggregate violent crime impacts of the interventions, accounting for spillovers <250m 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	-0.066	***0.048***	756	-49.6	-0.035	*0.393*	756	-26.7
Municipal services	-0.050	*0.225*	201	-10.0	-0.003	*0.964*	201	-0.7
Both					-0.122	***0.089***	75	-9.1
Subtotal				-59.6				-36.5
B. Spillover effects								
Intensive policing	0.001	*0.605*	52095	66.1	-0.004	*0.792*	52095	-193.1
Municipal services	-0.009	***0.075***	21286	-183.4	-0.014	***0.080***	21286	-305.1
Both					0.011	*0.319*	15772	165.8
Subtotal				-117.3				-332.4
Net increase (decrease) in crime				-176.8				-369.0
			95% CI	(-874, 378)			*95% CI*	(-1086, 287)
			90% CI	(-759, 280)			*90% CI*	(-955, 161)

Notes: This table presents the aggregate effect calculation for both interventions assuming spillovers <250m. 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.

Figure 7: Program impacts in nth percentile highest-crime street segments



Notes: We estimate equation 2 ten times, each time interacting each treatment indicator with an indicator for whether a segment is above the nth percentile of baseline crime levels among our experimental sample of segments, for n = 0, 10, 20, ... , 90. The coefficients on the treatment indicators indicate the effect on the higher crime segments above that percentile (hence the left side of the figure represents the highest crime "hot spots"). The figure graphs the direct treatment coefficients on intensive policing, municipal services, and the interaction, as well as the sum of these three coefficients.

the intervention period. The effect is imprecise, as the sample size drops dramatically. But these results are consistent with direct effects being roughly proportional to the levels of crime.

Note that in the two highest crime deciles, intensive policing alone is not associated with decreases in crime.[53] Any decrease is driven by municipal services (itself not statistically significant) or the combination of both (imprecise though the estimates may be). We conclude that our results are not being diluted by the inclusion of less hot segments.

Heterogeneity in spillover effects Above we estimated large adverse spillovers on average. Since many of the experimental streets have low to moderate levels of crime, logically the large adverse spillovers must come from the highest crime treated streets. Some of our results bear this expectation out. Estimating this heterogeneity in spillovers is complicated and noisy, however.

One reason is that the direct effects of policing and services alone are fairly small, and there is little heterogeneity in their direct treatment effects. Thus we expect the spillovers to be small and noisy too. A second reason is that the correct heterogeneity measure is not obvious. We prefer to use indicators for whether, for each spillover street, the maximum level of crime for any treated streets within 250 meters is in 0-25th, 25th-50th, 50th-75th,

53

or 75th-100th quartiles of baseline crime level for the experimental sample, while controlling for the expected value of that heterogeneity measure across all 1,000 randomizations.

We estimate heterogeneity in spillovers in Appendix C.7. We only see one large direct treatment effect, that from receiving both treatments. These treatment effects are largest for the both-treated segments that are also high crime. As expected, we estimate adverse spillovers from these highest-crime, both-treated streets. However, there are just 75 both-treated streets, and the highest crime ones are less than a quarter of these. So we are beginning to estimate spillover heterogeneity on impossibly small samples. Hence the noisiness of the estimates. The fact that we do not see large direct effects of policing or services alone makes it very difficult to say anything meaningful about spillover heterogeneity.

7.5 Robustness

Here we depart from the pre-specified models to consider other models and estimation methods. Qualitatively, our conclusions remain intact: there is weak evidence of a decrease in crime on directly treated streets, while signs continue to point to adverse (albeit imprecise) spillovers to nearby streets, especially for property crime.

Robustness to alternative spillover functions Table 10 reports the results of four alternative methods of spillover estimation. The first, in Panel A, replicates our earlier results from estimating equation (2). In Panel B, instead of an indicator S^P or S^M for any treated street within 250m, S^P and S^M are counts of the number of treated streets <250m. In Panel C, we estimate equation 3 using an exponential rate of decay rather than our fixed radii. Finally Panel D estimates the same equation with an inverse linear rate of decay. The coefficients on the two decay functions represent the expected increase in crimes as a segment moves a standard deviation closer to a treated segment.

Broadly speaking, we draw the same conclusions regardless of estimation methods. In terms of direct treatment effects, the results are similar to what we observed in Table 6 above: both intensive policing and municipal services have a negative but not statistically significant effect on crime when we ignore the interaction between treatments (not shown) but with the interaction term the reductions in crime are concentrated in the streets that received both treatments. There is no evidence of any crime reduction in streets that received just one intervention.

Turning to spillovers, in all cases intensive policing or the combination of both treatments (the sum of the three marginal effects) is associated with adverse spillovers on total crime, though these are not statistically significant . In all four cases, we also see the same difference between property and violent crimes as we did in the prespecified analysis. Intensive policing

Table 10: Estimated direct and spillover effects using alternative methods of spillover estimation, with RI p-values

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (2), (3), and (4)	Impact of spillovers <250m: Any intensive policing	Any municipal services	Both interventions	Sum of (6), (7), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Spillover indicator									
# of total crimes	0.283	0.035	0.072	-0.530	-0.422	0.019	0.006	-0.007	0.018
		0.664	*0.597*	***0.010***	***0.008***	*0.108*	*0.967*	*0.557*	*0.216*
# of violent crimes	0.103	-0.035	-0.003	-0.122	-0.160	-0.004	-0.014	0.011	-0.008
		0.395	*0.964*	***0.087***	***0.005***	*0.792*	***0.079***	*0.319*	*0.417*
# of property crimes	0.180	0.070	0.076	-0.408	-0.262	0.023	0.020	-0.017	0.025
		0.415	*0.556*	***0.029***	***0.060***	*0.041*	*0.316*	*0.203*	*0.045*
B. Spillover intensity									
# of total crimes	0.283	0.035	0.080	-0.411	-0.296	0.006	-0.003	0.000	0.003
		0.962	*0.656*	***0.079***	***0.040***	*0.375*	*0.598*	*0.964*	*0.727*
# of violent crimes	0.103	-0.014	0.005	-0.108	-0.116	-0.002	-0.002	0.003	-0.001
		0.584	*0.910*	*0.110*	***0.018***	*0.525*	*0.458*	*0.117*	*0.804*
# of property crimes	0.180	0.049	0.075	-0.303	-0.180	0.008	-0.001	-0.003	0.004
		0.866	*0.644*	*0.165*	*0.140*	*0.181*	*0.798*	*0.486*	*0.605*
C. Exponential decay									
# of total crimes	0.283	0.161	-0.044	-0.198	-0.082	0.087	-0.013	-0.025	0.048
		0.497	*0.604*	*0.542*	*0.540*	***0.060***	*0.614*	***0.090***	*0.447*
# of violent crimes	0.103	0.037	0.029	-0.149	-0.083	0.015	0.003	0.007	0.026
		0.711	*0.555*	***0.055***	*0.149*	*0.413*	*0.770*	*0.190*	*0.218*
# of property crimes	0.180	0.123	-0.074	-0.049	0.001	0.072	-0.017	-0.032	0.023
		0.548	*0.412*	*0.917*	*0.934*	***0.055***	*0.478*	***0.016***	*0.688*
D. Linear decay									
# of total crimes	0.283	0.137	0.004	-0.344	-0.203	0.027	-0.014	-0.003	0.010
		0.849	*0.860*	*0.166*	*0.205*	*0.588*	*0.282*	*0.750*	*0.901*
# of violent crimes	0.103	0.025	0.005	-0.119	-0.090	0.009	0.000	0.001	0.010
		0.826	*0.929*	*0.102*	*0.134*	*0.794*	*0.946*	*0.415*	*0.694*
# of property crimes	0.180	0.112	-0.001	-0.225	-0.113	0.018	-0.015	-0.004	0.000
		0.730	*0.817*	*0.387*	*0.436*	*0.585*	*0.195*	*0.477*	*0.732*

Notes: Randomization inference p-values are in italics. This table estimates the coefficients on spillovers, $\breve{\lambda}$, using equation 2 for panels A and B, and equation 3 for panels C and D. For panels A and B we estimate using both the experimental and nonexperimental streets. For panel B, in place of an indicator for any treated segment within a 250 radius, we use a count variable for the number of treated segments within 250m. In panels C and D, the weighted distance measures have been standardized to have zero mean and unit standard deviation.

tends to have larger adverse spillovers on property crimes than violent crimes. In panels A and C, the adverse spillover of intensive policing on property crime is statistically significant.

Robustness to ignoring different probabilities of treatment and the unusual patterns of clustering To see the effect of our design and estimation choices, in Appendix C.9 we estimate "naïve" treatment effects ignoring IPWs and randomization inference. Direct treatment effects are slightly smaller than in Table 7, but the patterns remain similar. The estimated spillover effects in this "naïve" case, however, are much larger and highly statistically significant compared to Table 7. Hence failing to account for interference between units and clustering of treatment conditions would have led us to severely exaggerate the degree to which policing pushes crime elsewhere.

7.6 Program impacts on state trust and legitimacy

We pre-specified three secondary outcomes capturing impacts on trust in and legitimacy of the state. First, an *opinion of police index* averaging 4 attitudes towards police: trust,; quality of work, overall satisfaction, and likelihood they would give information to police. Second, an *opinion of mayor index* that asks the same 4 questions for city government. Third, a crime reporting measure that captures the likelihood that people reported a crime to the police. This helps us understand whether administrative crime reporting changes with treatment, but is also a measure of collaboration and hence legitimacy.[54]

Overall, we see little evidence that the interventions increased trust in or legitimacy of the state. Table 11 reports ITT effects using equation 2. 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 (or even changes to a deterioration of the Mayor's opinion) when both treatments are received. This pattern is generally statistically significant at conventional levels. This heterogeneity across arms is hard to interpret and could reflect noise, so we are cautious and avoid drawing conclusions.

8 Discussion and conclusions

Police patrols and municipal services are the most elementary tools of city government. This study asks: what are the effects of these street-level bureaucrats on order and legitimacy in normal and high crime streets? Theoretically, both should increase the expected risk and

[54] In the state building and especially the counter insurgency literatures such civilian information, tips, and collaboration are among the chief indicators of state legitimacy.

Table 11: Impacts on state legitimacy allowing spillovers within 250m, with RI p-values

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (2), (3), and (4)	Impact of spillovers <250m: Any intensive policing	Any municipal services	Both interventions	Sum of (6), (7), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Opinion of police, z-score (+ better)	0.024	0.159	0.205	-0.327	0.037	-0.032	-0.003	0.199	0.165
		0.096	*0.080*	***0.009***	*0.868*	*0.512*	*0.542*	***0.065***	*0.510*
Opinion of mayor, z-score (+ better)	-0.014	0.022	0.177	-0.440	-0.241	-0.037	0.014	0.074	0.051
		0.859	***0.061***	***0.000***	***0.009***	*0.416*	*0.628*	*0.361*	*0.905*
Likeliness to report crime (0-3, + higher)	2.046	0.010	0.029	0.030	0.069	-0.006	0.013	0.050	0.057
		0.974	*0.685*	*0.574*	*0.312*	*0.716*	*0.928*	*0.332*	*0.416*

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 2, 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 three measures come from our citizen survey.

cost of committing crimes on the street segment, and so both interventions could be powerful tools of crime control. Bogotá offered an opportunity to evaluate these interventions on an unprecedented scale. And when we confine our analysis to the higher-crime streets, the results also speak to a larger literature on place-based interventions, including hot spots policing.

We show that crime in Bogotá was only moderately responsive to a street-level intensification of patrol time and intensified municipal services. These direct effects were large and significant only when they were jointly applied, but even then the total direct impacts were small. Moreover, the majority of this crime appears to shift to streets within 250m, especially property crime. The effects are proportionally larger in the highest crime streets, and so logically are the adverse spillovers. There is some indication that violent crimes are deterred more than property crimes, perhaps because they are more immediately opportunistic and do not have a sustained motive. Thus these findings bolster theories of crime deterrence that emphasize the importance of criminal motive.

Cost-effectiveness in this case is in the eye of the beholder. The city views the interventions as having little or no marginal cost, since they simply reallocated existing resources from some streets to others without raising their budgets or personnel. If so, then the main policy question is whether a moderately high probability of reducing murders and rapes by 8% is worth what seems to be a rise in property crime. This is a trade off that many police chiefs and mayors may reasonably make.

On the other hand, reallocating street-level bureaucrats had some unmeasured costs. There was a logistical cost of coordinating patrols, especially management time. It also made police patrols spend more time in unpleasant places. Officers told us they disliked the loss of autonomy and flexibility. There are also 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. The Mayor's office used scarce social and political capital to implement it. We believe one should measure this reform against the others it supplanted.

8.1 How do our results line up with existing evidence and meta-analyses?

At first glance, it might seem that the displacement of total crime to nearby streets runs against the U.S. literature. We have to compare with caution for several reasons. First, Bogotá and the US are different contexts. While Bogotá is a relatively rich and democratic society with a professionalized and relatively trusted police force, it is conceivable that poverty and inequality in Bogotá lead to more sustained motives for property crime (though this is highly speculative, especially when many US cities are also poor and unequal).

Second, as noted above, hot spots policing interventions vary in terms of intensity, concentration, crimes targeted, duration, and quality of approach. In Bogotá we study a much more routine increase in the intensity of normal services on a wider range of street segments.

A third difference is that our intervention focuses on direct effects and spillovers during the course of the intervention, and we look at spillovers over a larger-than-usual range of 250 meters. To the extent that spillovers are highest during the active phase of the interventions, or displace over larger spatial areas, there may be mechanical reasons for us to observe higher rates of displacement than studies that focus on a 2–block radius or post-intervention spillovers. Alternatively, past spillover bounds may have been too small and underestimated spillovers.

Finally, there are also reasons to question the consensus view that, on average, place-based interventions have positive spillovers (a diffusion of benefits). Two recent reviews argue that, on average, instances of positive spillovers outweigh negative ones across studies (Braga et al., 2012; Weisburd and Telep, 2016). Rather than question this aggregate mean, we think it is important to consider the confidence interval. Meta-analysis is extremely challenging here, since many of the component papers (especially early ones) do not report sufficient or comparable information. Also, only recently has it become more common for papers to move beyond simple t-tests of means to adjust standard errors for clustering of treatment assignment, or account for small sample distributions. Hence any aggregation involves either taking potentially problematic standard errors at face value, guesswork, or both.

But to use one meta-analysis as an example, of the 13 individual spillover estimates documented in Braga et al. (2014), 8 have a p-value smaller than 0.001.[55] Even in a study the size of Bogotá's, levels of precision were never nearly this high. Yet of these 8 $p < 0.001$ studies, the median number of target units including all treatment arms was below 30.

We have not redone the meta-analysis formally, and so we do not wish to make strong claims about what the aggregate confidence interval might be. Our point, rather, is that we we may want to regard the direction of spillovers as uncertain or unsettled. In addition, this illustrates the importance for future studies of data transparency, replicable evaluations, and ease of comparison across studies. And most of all it suggests that future meta-analysis should try to critically examine the confidence intervals in past studies.

[55] See Figure 2 in that paper.

8.2 Methodological lessons

As more urban policy experiments go to scale, we need practical tools and methods for dealing with the challenges that come from spillovers in dense interconnected networks. This isn't just important in cities, it is important for experiments in social networks and other settings where we worry about interference between units, and cannot experiment in separate and independent clusters.

Design-based approaches help in at least two ways. First, we show how multi-level randomization can be used to reduce the differential probabilities of assignment to spillover and control conditions that are so problematic for estimation. If nothing else, the design can ensure that the experiment has at least some proportion of control streets with a reasonable probability of assignment to all experimental conditions. Second, we show how design can estimate spillovers in a flexible way, with a minimum of ex ante assumptions. This flexibility is especially important when we don't have a strong sense of the structure of spillovers in advance. In Bogotá, we found evidence of spillovers in a catchment area considerably wider than the usual catchment area of 1–2 blocks. If true this could mean that the total spatial displacement is considerably greater than some papers estimate.

Besides illustrating the uses of design, this paper is also a rare practical example of the uses of randomization inference. RI has yet to gain currency in randomized control trials, in part because most times RI provides more or less the same conclusions as the usual clustered standard errors. Textbook cases for randomization inference, however, are when units of varying size have widely different probabilities of assignment to different experimental conditions, and when spillovers lead to fuzzy, difficult-to-model clustering. Large-scale urban interventions suffer from both problems, and we show how RI is a practical solution requiring relatively few assumptions. We expect RI to become an increasingly common approach to hypothesis testing in economics and criminology.

8.3 Lessons for place-based security interventions

Successful place-based interventions have two main characteristics. First, they have to improve security in the directly targeted areas. Second, they must minimize the chances of adverse spillovers. The first requirement is easier to satisfy than the second. However, our results point to strategies that could be more effective. One is that more intensive state presence, of both police and municipal services, seems to have had the largest direct effects on crime. It suggests there may be increasing returns to state presence. This combination deserves to be tested at scale, perhaps with additional or complementary strategies designed to prevent crime displacement.[56] But caution is warranted, since our estimates of displacement rose in tandem with direct treatment effects.

The Bogotá evidence also suggests that different kinds of crime might respond differently to interventions. Normal patrolling seems to have been most effective when targeted at segments with the most violent crime, or other crimes without sustained motives. This could include areas known for drunken brawls, confrontations between angry groups, or sexual assaults. Should the police want to avoid property crime displacement, it could mean a change in tactic, such as increasing arrests or seizures (which they did not do).

Similarly, it is possible that expanding targets beyond street segments could reduce displacement. A large literature has pushed attention to the level of street segments, corners and even addresses. But to the extent that hot spots cluster on nearby or adjacent streets, we may invite easy displacement by intervening and evaluating at the street segment level. It is possible that intervening in clusters would have larger direct effects and lower displacement of motivated crimes. This deserves testing. Clustering street segments can reduce statistical power, but there are ways to increase it back: by re-randomizing treatment at regular intervals, by increasing the intensity of treatment, or by experimenting in more than one municipality at once.

It is also worth noting that the broader policing literature has consistently found that more police are associated with lower crime (Levitt and Miles, 2006; Chalfin and McCrary, 2017b). Recent work by Chalfin and McCrary (2017a) suggests an especially large effect of aggregate police on violent crimes. One possibility is that a general increase in police per capita raises the probability of detection on every street and deters or captures even motivated criminals. This could be the key difference between intensive policing and greater manpower (though the latter is far more expensive). However, there may be ways to design place-based policing to minimize adverse spillovers. Treating hot spots at the level of a cluster or neighborhood rather than a street segment, as we mention above, is one such possibility.

In the end, our results bolster a view in criminology that the right question is not "do place-based policies deter crime?" but rather "what types of crime tend to be more responsive to state presence?" and "what tactics are suitable for deterring specific forms of crime?" Large-scale experiments offer an opportunity to test theories of crime, and the assumptions underlying interventions, rather than simply evaluate interventions themselves.

[56] Qualitatively, our interactions with the government and police patrols suggest other ways to increase direct impacts. One is less predictable policing, such as changing hot spots month to month. This has the advantage of increasing statistical power in an evaluation. Another is organizing hot spots in a more sophisticated manner, e.g. according to their risk at particular times of day or days of the week (such as schools at the start and end of the school day, or nightclubs in the evening).

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2016). Clustering as a Design Problem. *Working paper*.

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.

Aronow, P. M. and C. Samii (2013, May). Estimating Average Causal Effects Under General Interference, with Application to a Social Network Experiment. *Working paper*.

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.

Blanes I Vidal, J. and G. Mastrobuoni (2017). Police Patrols and Crime. *Centre for Economic Policy Research Working Paper DP12266.*

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.

Braga, A. A., A. V. Papachristos, and D. M. Hureau (2014). The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice Quarterly 31*(4), 633–663.

Braga, A. A., B. C. Welsh, and C. Schnell (2015). Can policing disorder reduce crime? A systematic review and meta-analysis. *Journal of Research in Crime and Delinquency 52*(4), 567–588.

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.

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

Chalfin, A. and J. McCrary (2017b). Criminal Deterrence: A Review of the Literature. *Journal of Economic Literature 55*(1), 5–48.

Chiba, S. and K. Leong (2014). Behavioral economics of crime rates and punishment levels. *Working Paper*.

Clarke, R. V. and D. Weisburd (1994). Diffusion of crime control benefits: Observations on the reverse of displacement. *Crime prevention studies 2*, 165–184.

Collazos, D., E. Garcia, D. Mejia, D. Ortega, and S. Tobon (2018). Hotspots policing in a high crime environment: An experimental evaluation in Medellin. *In progress*.

Dell, M. (2015). Trafficking networks and the Mexican drug war. *American Economic Review 105*(6), 1738–79.

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.

Ehrlich, I. (1973). Participation in Illegitimate Activities: A Theoretical and Empirical Investigation. *The Journal of Political Economy 81*, 521–565. 3.

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

Ferraz, C., J. Monteiro, and B. Ottoni (2016). State Presence and Urban Violence: Evidence from Rio de Janeiro's Favelas. *Working Paper*.

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

Guerette, R. and K. Bowers (2009). Assessing the extent of crime displacement and diffusion of benefits: a review of situational and crime prevention evaluatoins. *Criminology*.

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.

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

Levitt, S. D. and T. J. Miles (2006). Economic Contributions to the Understanding of Crime. *Annual Review of Law and Social Science 2*(1), 147–164.

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.

Muggah, R., I. S. D. Carvalho, N. Alvarado, L. Marmolejo, and R. Wang (2016). Making Cities Safer: Citizen Security Innovations. *Igarapé Institute, Inter-American Development Bank, World Economic Forum Strategic*(June).

Nussio, E. and E. Norza Cespedes (2018). Deterring delinquents with information. evidence from a randomized poster campaign in Bogota. *PLOS ONE 13*(7), 1–20.

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.

Sherman, L., M. Buerger, and P. Gartin (1989). beyond dial-a-cop: a randomized test of repeat call policing (recap). In *Brryond Crmie and Punishment*. 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.

Vazquez-Bare, G. (2017). Identification and Estimation of Spillover Effects in Randomized Experiments. *arXiv:1711.02745 [econ]*.

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. M. 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.

Weisburd, D. L., L. a. Wyckoff, J. E. Eck, and J. Hinkle (2006). Does Crime Just Move Around the Corner? A Study of Displacement and Diffusion in Jersey City, NJ. *Criminology 44*(August), 549–592.

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

Appendix for online publication

Contents

A Statistical power analysis

Figure A.1 takes studies from recent systematic reviews and plot sample size and effect sizes for both direct and spillover effects.[57] One major takeaway is that most studies are not ex ante powered to detect the average direct effect across studies, of 0.17 standard deviations. The figure displays statistical power curves, representing the minimum effect size that we would expect to be able to detect with 80% confidence. While covariate adjustment and blocking strategies could improve statistical power slightly, these would produce at best marginal gains in precision.[58] Note that even the largest studies do not exceed 50 or 100 treated hot spots, with a similarly modest number of spillover segments.[59] The average effect

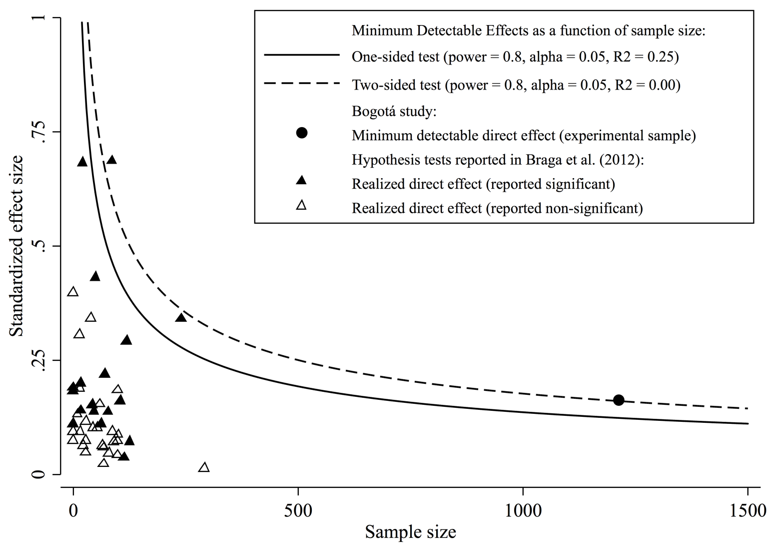
[57]This is one reason why most studies were designed to address direct treatment effects, and spillovers are a secondary outcome. One exception is Weisburd et al. (2006), who study drug and prostitution hot spots. Their findings suggest the benefits from the intervention diffuse to nearby areas.

[58]We generate the power curves assuming simple randomization and treatment assignment for half of the experimental sample. 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. 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 shown in Braga et al. (2014), 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.

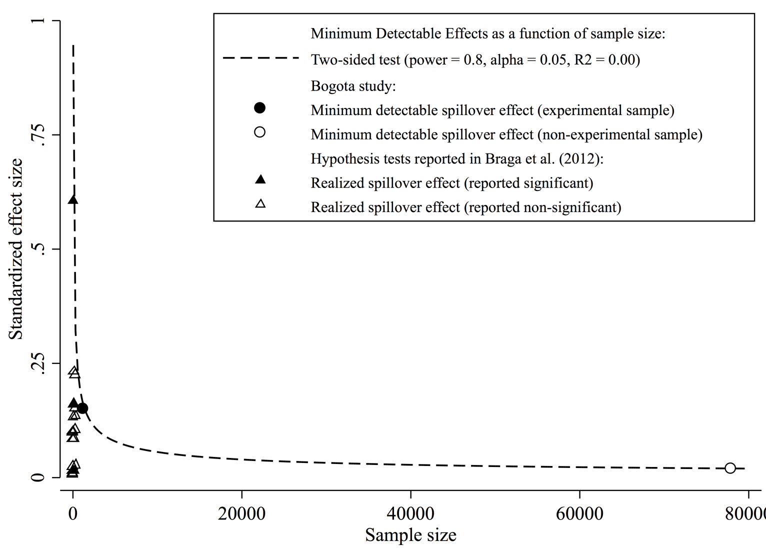
[59]Randomized controlled trials of intensive policing have sample sizes of 110 hot spots (55 treated) in Minneapolis (Sherman and Weisburd, 1995), 56 hot spots (28 treated) in Jersey City Weisburd and Green

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 figures depict minimum detectable effects and realized effect sizes as a function of sample size and the presence of other explanatory variables (via R-squared). 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. Triangles represent a hypothesis test from previous studies and circles represent the minimum detectable effects in our study.

size for direct hot spots treatment across the studies is 0.17 standard deviations, and 0.24 if statistically significant. We only report MDEs for studies for which it was possible to do so with the information in published papers, however. 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.

(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. (2018). 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.

B Additional data and design details

B.1 Patrolling time

Figure B.1 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.[60] 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.[61]

B.2 Inverse probability weighting

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.

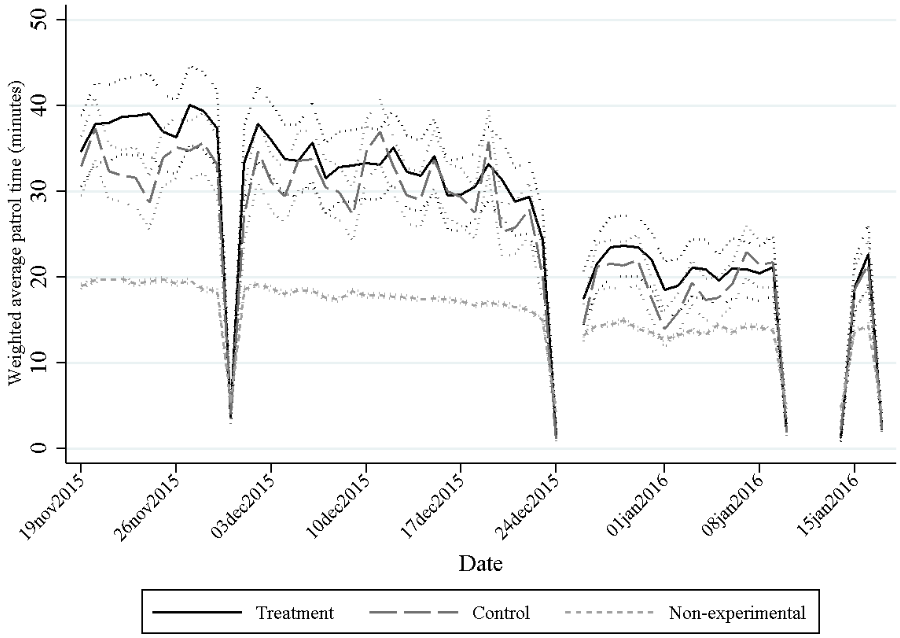
Figure B.2 compares two maps. The first map displays the number of baseline administrative crimes between 2012 and 2015 for each segment, while the second one displays each segment's probability of being within 250m of hotspots receiving hotspot policing and municipal services (based on 1,000 randomizations). In areas with lots of crime, non-experimental

[60] 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 40m 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. 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.

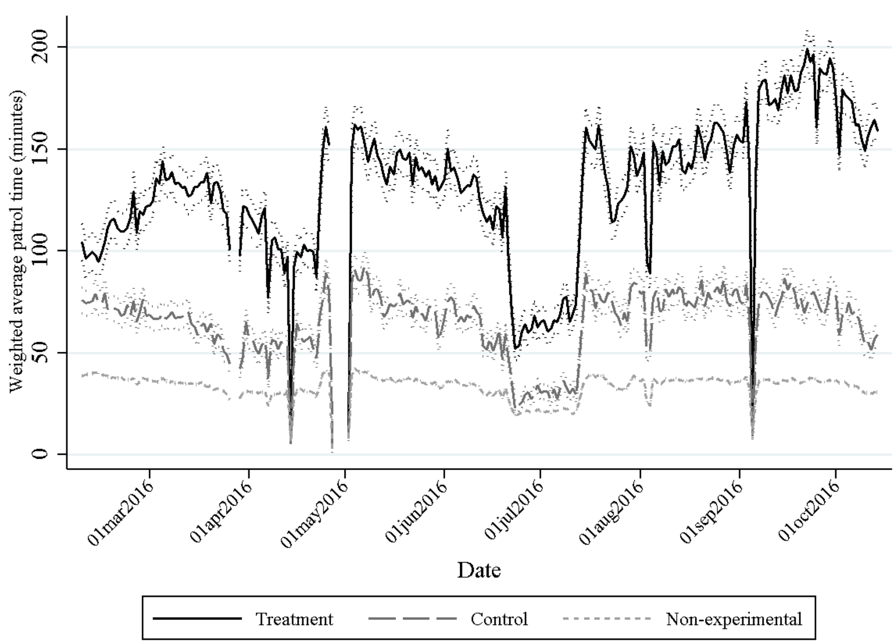
[61] For our estimates, we follow each GPS device chronologically, thus we track the moment at which the device enters a street and when does it leave. We made two assumptions to estimate patrolling time: (i) If we see only one GPS ping in a street and then the device moves to other streets, we impute 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).

Figure B.1: Evolution of patrolling time in the pre-treatment and treatment periods

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

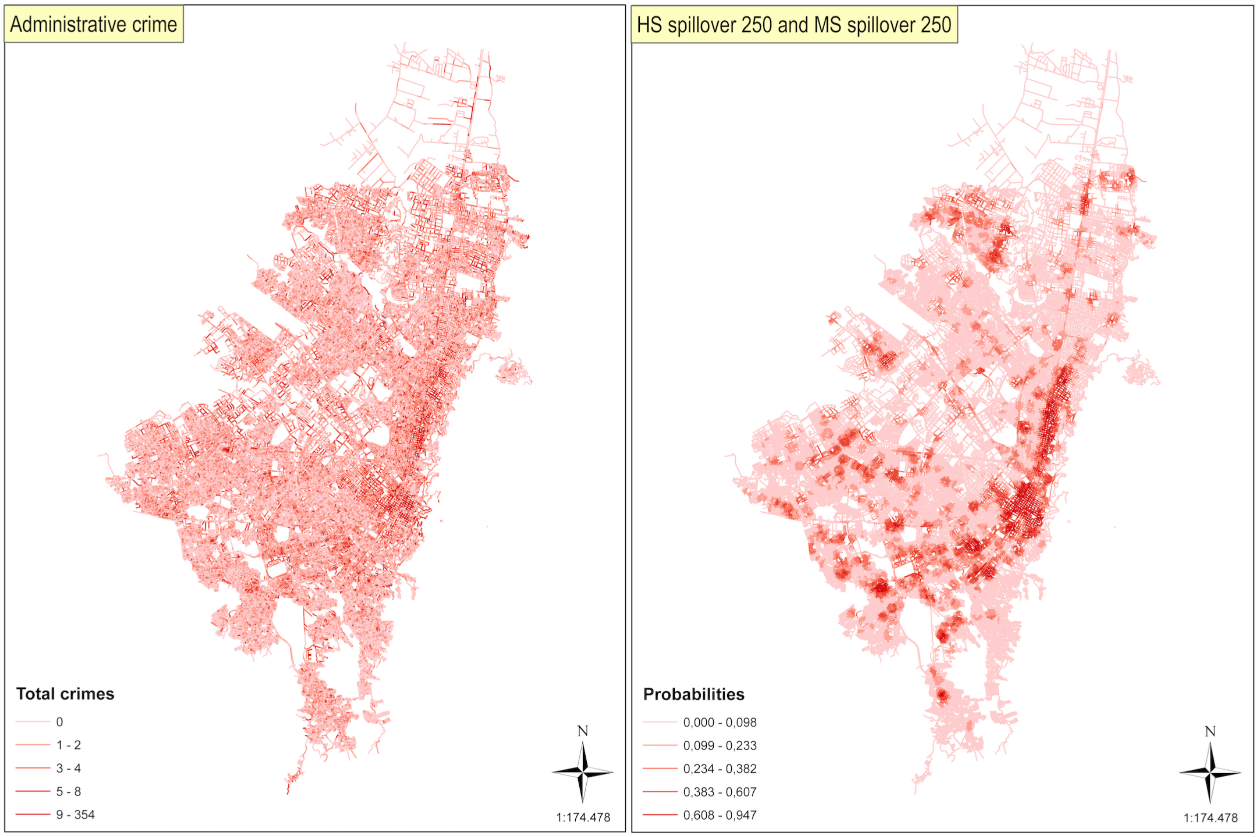


(b) Treatment period (February 2016 – October 2016)



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.

Figure B.2: Maps of baseline crime and probability of being spillover <250m 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.

units have a higher probability of being a <250m 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 <250 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).

In table B.2 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.

Table B.2: IPW bias

		Experimental sample							
		Treatment effect				Spillover effect			
Outcome	Interaction included? (1)	Intensive policing (2)	Municipal services (3)	Interaction effect (4)	Both (2+3+4) (5)	Intensive policing (6)	Municipal services (7)	Interaction effect (8)	Both (6+7+8) (9)
Insecurity index, z-score (+ more insecure)	No	0.040	0.051		0.091	0.067	0.084		0.151
	Yes	0.027	0.038	0.031	0.095	0.083	0.096	-0.039	0.140
Perceived risk index, z-score (+ riskier)	No	0.037	0.039		0.076	0.064	0.072		0.135
	Yes	0.024	0.026	0.031	0.081	0.078	0.083	-0.036	0.125
Crime index, z-score (+ more crime)	No	0.030	0.045		0.075	0.049	0.068		0.117
	Yes	0.021	0.037	0.020	0.078	0.061	0.078	-0.030	0.108
Perceived & actual incidence of crime, z-score (survey)	No	0.035	0.040		0.075	0.062	0.061		0.123
	Yes	0.021	0.026	0.032	0.080	0.078	0.073	-0.040	0.111
# crimes reported to police on street segment (admin)	No	0.013	0.044		0.057	0.011	0.065		0.077
	Yes	0.017	0.048	-0.009	0.055	0.012	0.067	-0.003	0.076

		Non-experimental sample			
		Spillover effect			
	Interaction included? (1)	Intensive policing (6)	Municipal services (7)	Interaction effect (8)	Both (6+7+8) (9)
Insecurity index, z-score (+ more insecure)	No	0.003	0.010		0.013
	Yes	0.002	0.009	0.001	0.012
Perceived risk index, z-score (+ riskier)	No	0.002	-0.006		-0.004
	Yes	-0.003	-0.012	0.010	-0.005
Crime index, z-score (+ more crime)	No	0.003	0.023		0.026
	Yes	0.007	0.027	-0.008	0.026
Perceived & actual incidence of crime, z-score (survey)	No	-0.001	0.009		0.008
	Yes	0.004	0.014	-0.010	0.008
# crimes reported to police on street segment (admin)	No	0.009	0.045		0.054
	Yes	0.011	0.045	-0.003	0.054

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.

B.3 Departures from the pre-analysis plan

The estimation procedure used in this paper is slightly different from the ones we described in our pre-analysis plan (PAP).[62] 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 pre-analysis plan says we would run the following WLS regression:

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

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 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 B.3.1, 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 pre-analysis plan so that our empirical strategy is compatible with the realities of our factorial design. First, we can base our probability weights on 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:

$$Y_{sqp} = \beta_0 + \theta_H * T_H + \theta_M * T_M + \theta_H * S^M + \emptyset * X_{sqp} + \gamma_p + \varepsilon_{sqp} \tag{5}$$

[62] https://www.socialscienceregistry.org/trials/1156.

Table B.3.1: 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	0.37	1.40	1.74
assignment	>250m	458	0.11	0.13	0.62
		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.

Including an additional indicator for being a hot spot policing spillover in this regression allows us to estimate all four effects (direct effect of hot spot 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 (1) in the main paper where $\beta_3 = 0$. Thus the regressions used in this paper correctly estimate 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 B.3.2 displays the hot spots policing effect while table displays the municipal services effects. Meanwhile, table B.3.4 displays the interaction effects. Most of the differences for the treatment effects are coming from the use of different weights. In Table B.3.4 (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.

Table B.3.2: Hot spots policing impacts on insecurity, pre-specified regressions

	ITT of assignment to:				
	Accounting for 250-500m spillovers	Accounting for spillovers <250m			No spillovers
Dependent variable	HSP outer spillover	HSP treated	HSP inner spillover (experi-mental)	HSP inner spillover (non-experimental)	HSP treated
	(1)	(2)	(3)	(4)	(5)
Insecurity index, z-score (+ more insecure)	0.078	-0.106	-0.055	0.122	-0.063
	0.366	*0.291*	*0.816*	*0.223*	*0.193*
Perceived risk index, z-score (+ riskier)	-0.014	-0.110	-0.080	0.120	-0.067
	0.834	*0.232*	*0.515*	*0.303*	*0.151*
Crime index, z-score (+ more crime)	0.144	-0.067	-0.011	0.083	-0.039
	0.169	*0.520*	*0.903*	*0.300*	*0.445*
Perceived & actual incidence of crime, z-score	0.175	-0.087	-0.030	0.110	-0.045
	0.137	*0.355*	*0.884*	*0.330*	*0.346*
# crimes reported to police on street segment	0.050	-0.011	0.029	0.017	-0.016
	0.651	*0.943*	*0.760*	***0.035***	*0.861*

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.

Table B.3.3: Municipal services impacts on insecurity, pre-specified regressions

	ITT of assignment to:				
	Accounting for 250-500m spillovers	Accounting for spillovers <250			No spillovers
Dependent variable	MS outer spillover	MS treated	MS inner spillover (experi-mental)	MS inner spillover (non-experimental)	MS treated
	(1)	(2)	(3)	(4)	(5)
Insecurity index, z-score (+ more insecure)	-0.097	-0.138	0.082	-0.092	-0.152
	0.471	*0.092*	*0.173*	*0.419*	***0.005***
Perceived risk index, z-score (+ riskier)	-0.020	-0.137	0.031	-0.144	-0.128
	0.879	*0.101*	*0.507*	*0.189*	***0.008***
Crime index, z-score (+ more crime)	-0.142	-0.092	0.106	-0.009	-0.126
	0.161	*0.267*	*0.102*	*0.917*	***0.018***
Perceived & actual incidence of crime, z-score	-0.060	-0.129	0.056	0.032	-0.129
	0.656	*0.149*	*0.355*	*0.758*	***0.023***
# crimes reported to police on street segment	-0.264	0.006	0.175	-0.037	-0.088
	0.057	*0.971*	*0.141*	***0.000***	*0.379*

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 B.3.4: Interaction impacts on insecurity, pre-specified regressions

	ITT of assignment to:					
	Accounting for spillovers <250m			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)	-0.177	-0.179	0.058	-0.077	-0.074	-0.185
	0.152	*0.140*	*0.572*	*0.220*	*0.274*	*0.120*
Perceived risk index, z-score (+ riskier)	-0.191	-0.183	0.058	-0.087	-0.053	-0.139
	0.120	*0.112*	*0.567*	*0.164*	*0.414*	*0.226*
Crime index, z-score (+ more crime)	-0.104	-0.115	0.038	-0.042	-0.070	-0.170
	0.368	*0.360*	*0.706*	*0.528*	*0.298*	*0.179*
Perceived & actual incidence of crime,	-0.147	-0.255	0.229	-0.057	-0.138	0.040
z-score	*0.250*	*0.047*	*0.137*	*0.405*	*0.052*	*0.687*
# crimes reported to police on street	-0.003	0.182	-0.329	-0.004	0.076	-0.526
segment	*0.940*	*0.372*	*0.295*	*0.957*	*0.579*	*0.024*

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.

C Additional results and robustness analysis

C.1 Summary statistics

Our primary outcomes are two insecurity measures: perceived risk and crime incidence. Table C.1 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.

C.2 Tests of spillovers

Table C.2 reports the p-values from our preferred, general test of spillovers. It takes the means for the 4×5 experimental conditions in Table 2 in the paper and tests for differences between pairs of columns (for municipal services) and pairs of rows (for intensive policing). Using our pre-specified threshold of p<0.1, we observe statistically significant spillovers with 250m for municipal services, but not in the 250-500m region. For intensive policing, however, none of the p-values are below 0.1. We see some indication of <250m spillovers from municipal services in one of the two outcomes (crime incidence), but spillovers are not statistically significant in the large non-experimental sample.

This is one reason why we see more statistically significant spillovers in Table 6. We should also have addressed how we would treat economically large spillovers around or below $p = 0.1$. Because the spillovers in Table C.2 are weak, there is a reasonable argument for calculating treatment effects ignoring spillovers.

C.3 Program impacts on officially reported crime, un-pooling the experimental and experimental samples

Table C.3 replicates Table 6 in the main paper, except it estimates equation (1) on the experimental sample alone instead of equation (2) on the pooled sample of experimental and nonexperimental units. The nonexperimental spillovers are then esitmated separately. Conclusions do not change materially.

C.4 Back-of-the-envelope calculation of the effects of scaling the dual treatment and no treatment

In Table C.4 we calculate the direct and indirect effects assuming that the 882 street segments that received either the intensive policing or the municipal services treatment received both treatments. In order to obtain the direct treatment effect, we add the intensive policing and the municipal services effects, and subtract the interaction effect. The estimated impact is the result of the direct treatment effect times the number of units that received either the intensive policing, the municipal services, or both treatments. Similarly, to calculate the indirect treatment effect for the experimental and non-experimental sample, we add the

Table C.1: Summary statistics for the primary security outcomes, all experimental conditions (N = 1,919 hot spots)

		Municipal services assignment				
		Treated	<250m	250-500m	>500m	Ineligible
		(1)	(2)	(3)	(4)	(5)
A: Perceived risk (z-score)						
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)						
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.

Table C.2: 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 250-500m 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.

treatment coefficients and subtract the interaction coefficient. Then, we multiply each of these effects by the number of street segments that were within <250m spillover region of a treated unit by any or both treatments.

C.5 Program impacts including survey-based insecurity measures, no interaction term

When examining officially reported crimes, the main paper reported estimates of treatment and spillover effects with and without the interaction between intwensive policing and municipal services. However, the paper only reported results with the interaction terms when considering our broader measures of insecurity weith survey data. Table C.5 reports these results. There is no material difference in conclusions without the interaciton term.

C.6 Aggregate effects for crime subgroups

In Tables C.6.1 and C.6.2, we display the aggregate effects on additional crime subgroups: property crime, as well as homicides and rapes only.

Table C.3: Estimated aggregate impacts of the interventions, accounting for spillovers within <250m

	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	7.1
Municipal services	-0.076	*0.783*	201	-15.2	0.089	*0.367*	201	17.9
Both					-0.437	***0.043***	75	-32.8
Subtotal				-86.0				-7.8
B. Spillover, experimental sample								
Intensive policing	0.061	*0.595*	705	42.7	0.143	*0.315*	705	100.7
Municipal services	0.176	***0.056***	546	96.3	0.255	***0.025***	546	139.1
Both					-0.272	*0.196*	281	-76.5
Subtotal				138.9				163.3
C. Spillover, non-experimental sample								
Intensive policing	0.016	*0.101*	51390	844.7	0.013	*0.205*	51390	677.4
Municipal services	-0.003	*0.394*	20740	-65.8	-0.006	*0.484*	20740	-124.9
Both					0.005	*0.973*	15,491	85.0
Subtotal				778.9				637.5
Net increase (decrease) in crime				831.9				793.0
			95% CI	(-780, 2054)			*95% CI*	(-1001, 2199)
			90% CI	(-434, 1848)			*90% CI*	(-689, 1989)

Notes: 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.4: Estimated aggregate impacts of the interventions, accounting for spillovers within <250m

	Dependent variable: # of crimes reported to police on segment (administrative data)		
	Interaction between treatments		
Impacts of treatment	Coeff. (1)	# segments (2)	Estimated total impact = (1) × (2) (3)
A. Direct treatment effect			
Intensive policing + Municipal services	-0.423	882	-373.1
B. Indirect effects			
Intensive policing + Municipal services	0.018	57609	1,037.0
Net increase (decrease) in crime			663.9

Notes: 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.5: Program impacts on security in the experimental sample, accounting for spillovers within 250m, with p-values from randomization inference, no interaction between treatments (N=2,396)

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	ITT of assignment to: Any municipal services	Impact of spillovers <250m: Any intensive policing	Impact of spillovers <250m: Any municipal services
	(1)	(2)	(3)	(4)	(5)
Insecurity index, z-score (+ more insecure)	0.005	-0.129	-0.153	-0.011	0.060
		0.274	*0.172*	*0.626*	*0.137*
Perceived risk index, z-score (+ riskier)	0.049	-0.139	-0.127	-0.041	0.006
		0.168	*0.184*	*0.873*	*0.470*
Crime index, z-score (+ more crime)	-0.040	-0.076	-0.128	0.023	0.094
		0.520	*0.274*	*0.448*	***0.043***
Perceived & actual incidence of crime, z-score (survey)	0.059	-0.056	-0.132	0.004	0.051
		0.750	*0.204*	*0.590*	*0.200*
Crime data index, z-score (admin)	-0.117	-0.102	0.005	0.054	0.151
		0.241	*0.665*	*0.571*	***0.056***
# crimes reported to police on street segment (admin)	0.743	-0.094	-0.088	0.061	0.176
		0.488	*0.687*	*0.595*	***0.056***

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 and 3) and the spillover effects onto the control hot spots only in the experimental sample only (Columns 3 and 5) via a WLS regression of each outcome on treatment indicators, 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.6.1: Estimated aggregate property crime impacts of the interventions, accounting for spillovers <250m 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	-0.033	*0.812*	756	-24.8	0.071	*0.417*	756	53.5
Municipal services	-0.084	*0.322*	201	-16.8	0.076	*0.558*	201	15.3
Both					-0.408	***0.029***	75	-30.6
Subtotal				-41.7				38.1
B. Spillover, experimental sample								
Intensive policing	0.015	*0.085*	52095	805.6	0.023	*0.041*	52095	1198.7
Municipal services	0.011	***0.874***	21286	225.8	0.020	***0.317***	21286	424.4
Both					-0.018	*0.204*	15772	-277.2
Subtotal				1031.4				1345.9
Total				989.7				1384.0
			95% CI	(-438, 1813)			*95% CI*	(-122, 2681)
			90% CI	(-282, 1646)			*90% CI*	(153, 2453)

Notes: This table presents the aggregate effect calculation for both interventions assuming spillovers <250m. 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.6.2: Estimated aggregate homicide and rape impacts of the interventions, accounting for spillovers <250m 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			
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	-0.011	*0.230*	756	-8.6	-0.008	*0.468*	756	-5.8
Municipal services	-0.006	*0.640*	201	-1.2	0.000	*0.942*	201	0.0
Both					-0.016	*0.375*	75	-1.2
Subtotal				-9.8				-7.0
B. Spillover, experimental sample								
Intensive policing	0.006	*0.642*	705	4.1	0.013	*0.358*	705	9.0
Municipal services	-0.009	*0.352*	546	-5.0	-0.002	*0.875*	546	-1.2
Both					-0.022	*0.233*	281	-6.3
Subtotal				-0.8				1.5
C. Spillover, non-experimental sample								
Intensive policing	-0.001	*0.418*	51190	-56.4	-0.002	*0.129*	51190	-107.1
Municipal services	0.000	*0.871*	20740	6.2	-0.001	*0.444*	20740	-17.0
Both					0.002	*0.324*	15491	31.9
Subtotal				-50.2				-92.1
Net increase (decrease) in crime				-60.8				-97.6
			95% CI	(-179, 55)			*95% CI*	(-233, 33)
			90% CI	(-165, 40)			*90% CI*	(-210, 15)

Notes: This table presents the aggregate effect calculation for both interventions assuming spillovers <250m. 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.7 Spillover heterogeneity by baseline crime

Table C.7 shows the results of a specification similar to Equation (2) where the each treatment and spillover effect is replaced with 4 effects, one for each quartile of baseline crime level. For the treated streets, estimating differing effect is done through simple interaction terms. We found the 1st, 2nd, 3rd, and 4th quartile of baseline crime level in experimental streets. Then, we constructed 4 dummy variables for all streets in the city based on whether their crime level fell into the 1st, 2nd 3rd, or 4th quartile. That is to say, the mean of each quartile dummy variable was .25 within the experimental sample of 1919 streets, but not for the entire regression sample of 79,000 streets. To construct spillover dummies from these quartiles, we grouped each spillover street based on the maximum crime level of treated streets nearby. If a spillover street had all their nearby treated streets in the first quartile, then they received a 1 for their first-quartile spillover dummy, and zero for the others. However if one those treated streets within 250 meters fell into the second quartile, they would receive only a 1 for the second-quartile spillover dummy, and zero for the others. Note that though the experimental sample was by definition evenly divided into 4 quartiles, treatment assignment was actually biased towards lower crime streets. This is due to the restriction of only selecting two streets from each police quadrent into the treatment.

Results are noisy, but consistent with both the main analysis and those in Figure 7 in the main paper. In particular, the combined effect for the combination of both direct treatments is highest in streets with high levels of baseline crime. However any conclusions must be tempered by the fact that estimates for the effect of combined treatment at the top quartile are derived from a treated sample of only 14 steets.

C.8 Marginal effects of extra patrolling time

We estimate the results instrumenting patrolling time (measured in hours) with treatment assignment to intensive policing. We also explore if the marginal effects of additional patrolling time differ over varying levels of baseline crime. Table C.8 reports these results. Standard errors are clustered at the unit of randomization. For estimating instrumental variables results, we cannot use randomization inference to estimate exact p-values, as we would need to know how would the compliance levels be under each possible randomization. This implies we are over-stating precision in these analyses. Additionally, because we are unable to use randomization inference to correct for clustering of spillovers, columns (4)-(7) report a regression where we exclude streets in the experimental sample that are withing 250m of a treated street.

Results from both tables are similar, hence we focus on the no-spillovers case (Table C.8). Column (1) presents the OLS results. Note that, since patrolling time is endogenous to crime levels, the coefficient is positive. Column (2) presents the instrumental variables estimates. In this case, the sign of the coefficient of patrolling time is reversed, as expected, and suggests a negative relationship between patrolling time and the number of reported crimes. Column (3) includes an interaction of patrolling time and baseline crime. Note the marginal effect of one additional hour of patrolling time is of about 0.13 fewer crimes, but this effect is decreasing as the baseline crime levels are larger (see the positive sign of the

Table C.7: Direct and spillover effects of the program by initial levels of crime, using mutually exclusive baseline crime quartiles (p-values in italics)

	Direct effects for treated streets in each quartile of crime level relative to the crime distribution of the experimental sample				Spillover effects for streets where the treated street within 250m is at maximum in each quartile of crime			
Condition	0-25%	25-50%	50-75%	75-100%	0-25%	25-50%	50-75%	75-100%
Intensive policing	-0.0601	0.1384	0.2136	-0.1596	-0.0062	0.0455	0.0060	-0.0450
	0.33	*0.37*	*0.24*	*0.89*	*0.57*	***0.01***	*0.66*	*0.53*
Municipal services	0.1334	0.0022	0.4132	0.0082	-0.0034	0.0011	0.0114	0.0732
	0.43	*0.77*	***0.06***	*0.74*	*0.78*	*0.94*	*0.90*	*0.38*
Interaction	-0.1403	-0.2456	-0.5369	-1.7170	0.0007	-0.0025	-0.0622	0.0300
	0.71	*0.29*	*0.18*	***0.03***	*0.87*	*0.93*	***0.10***	*0.90*
Both (sum)	-0.1799	-0.1051	0.0899	-1.8684	-0.0089	0.0441	-0.0442	0.0582
	0.78	*0.37*	*0.79*	***0.03***	*0.65*	*0.12*	*0.33*	*0.65*
	Number of street segments in each direct treatment condition				Number of streets in each spillover condition			
Condition	0-25%	25-50%	50-75%	75-100%	0-25%	25-50%	50-75%	75-100%
Intensive policing	221	209	196	130	17610	16912	14045	8743
Municipal services	55	56	45	45	6694	6539	4315	3795
Both	27	19	15	14	3792	2762	1650	1421
	Mean number of crimes at endline in each direct treatment condition				Mean number of crimes at endline in each spillover condition			
Condition	0-25%	25-50%	50-75%	75-100%	0-25%	25-50%	50-75%	75-100%
Intensive policing	0.24	0.62	1.13	2.33	0.21	0.29	0.37	0.68
Municipal services	0.24	0.61	1.29	1.96	0.22	0.31	0.45	0.85
Both	0.19	0.53	1.27	1.21	0.24	0.29	0.45	0.75

Table C.8: Instrumental variables results (full sample)

Dependent variable	All experimental streets			Excluding control hotspots within 250m of treated hotspots		
		Instrumental variables			Instrumental variables	
	OLS	No interaction	Interaction with baseline crime	OLS	No interaction	Interaction with baseline crime
	(1)	(2)	(3)	(4)	(5)	(6)
A. IV Results. Dependent variable is # of total reported crimes						
Patrolling time (hours)	0.012	-0.122	-0.133*	-0.003	-0.057	-0.075
	[0.033]	[0.077]	[0.069]	[0.032]	[0.074]	[0.074]
Patrolling time (hours) × baseline crime			0.022			0.058
			[0.073]			[0.065]
B. First stage results. Dependent variable is patrolling time (hours)						
Assigned to HS treatment		1.277***	1.264***		1.402***	1.378***
		[0.074]	[0.076]		[0.091]	[0.092]
HS treatment × baseline crime			0.02			0.092
			[0.055]			[0.058]
C. Summary statistics for each regression						
Observations	1,919	1,919	1,919	1,214	1,214	1,214
Weighted Avg. # of reported crimes	1.038	1.038	1.038	0.890	0.890	0.890
Weighted Avg. patrolling time (hours)	2.163	2.163	2.163	2.220	2.220	2.220

Notes: This table reports average treatment effects on the treated (ATT) estimates of the effects of intensive policing, via a weighted instrumental variables regressions of reported crimes on patrolling time (in hours) instrumented with treatment assignment (or the interactions instrumented accordingly). Regressions also include police station (block) fixed effects, and baseline covariates (and the relevant exogenous regressions accordingly). 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. * significant at the 10 percent, ** significant at the 5 percent, *** significant at the 1 percent.

interaction).

C.9 Impacts without re-weighting and randomization inference

Table C.9 reproduces Table from the paper, but without IPWs and randomization inference. The direct treatment effects are generally smaller but the patterns are still similar. However, the spillover effects in these results are huge (.18 standard deviations for hot spots policing, 0.31 standard deviations for municipal services). 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.

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.

Table C.9: Naïve treatment effects

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interven-tions	Sum of (2), (3), and (4)	Impact of spillovers <250: Any intensive policing	Any municipal services	Both interven-tions	Sum of (6), (6), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Insecurity index, z-score (+ more insecure)	0.066	0.006	-0.069	-0.090	-0.153	0.180	0.308	-0.225	0.262
		[.057]	[.091]	[.130]	[.102]	[.064]***	[.066]***	[.096]**	[.078]***
Perceived risk index, z-score (+ riskier)	-0.018	-0.022	-0.080	-0.096	-0.198	0.107	0.215	-0.193	0.129
		[.058]	[.088]	[.131]	[.104]*	[.063]*	[.064]***	[.088]**	[.076]*
Crime index, z-score (+ more crime)	0.128	0.031	-0.035	-0.053	-0.058	0.192	0.297	-0.182	0.308
		[.055]	[.095]	[.134]	[.103]	[.067]***	[.067]***	[.105]*	[.084]***
Perceived & actual incidence of crime, z-score	0.002	-0.002	-0.120	0.101	-0.021	0.143	0.240	-0.211	0.172
		[.062]	[.090]	[.143]	[.120]	[.067]**	[.070]***	[.097]**	[.082]**
# crimes reported to police on street segment	1.334	0.086	0.135	-0.333	-0.113	0.238	0.334	-0.082	0.491
		[.091]	[.160]	[.217]	[.164]	[.124]*	[.115]***	[.197]	[.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 (1). 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.